Analysis of Glycopeptides by HPLC MS and Detection of Leukemic Cells by Microfluidics using Time Delayed Integration Spectral Flow Cytometry

By
Wenting Hu

Submitted to the graduate degree program in the Department of Chemistry and the Graduate Faculty of the University of Kansas in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

_____
Chairperson: Dr. Steven A. Soper

_____
Co-chair: Dr. Heather Desaire

_____
Dr. Susan Lunte

_____
Dr. Davis Weis

_____
Dr. Mikhail V. Barybin

_____
Dr. Brandon Dekosky

Date Defended: Jan 10, 2019

The dissertation committee for Wenting Hu certifies that this is the approved version of the following dissertation:

Analysis of Glycopeptides by HPLC MS and Detection of Leukemic Cells by Microfluidics using Time Delayed Integration Spectral Flow Cytometry

_____

Chairperson: Dr. Steven A. Soper

_____

Co-chair: Dr. Heather Desaire

Date approved: Jan 10, 2019

ABSTRACT

Protein glycosylation drives many biological processes and serves as marker for disease; therefore, the development of tools to study glycosylation is an essential and growing area of research. Mass spectrometry can be used to identify both the glycans of interest and the glycosylation sites to which those glycans are attached, when proteins are proteolytically digested and their glycopeptides are analyzed by a combination of high-resolution mass spectrometry (MS) and tandem mass spectrometry (MS/MS) methods. One major challenge in these experiments is collecting the requisite MS/MS data. The digested glycopeptides are often present in complex mixtures and in low abundance, and the most commonly used approach to collect MS/MS data on these species is data-dependent acquisition (DDA), where only the most intense precursor ions trigger MS/MS. DDA results in limited glycopeptide coverage. Semi-targeted data acquisition is an alternative experimental approach that can alleviate this difficulty. However, due to the massive heterogeneity of glycopeptides, it is not obvious how to expediently generate inclusion lists for these types of analyses. To solve this problem, we developed the software tool GlycoPep MassList, which can be used to generate inclusion lists for liquid chromatography tandem-mass spectrometry (LC-MS/MS) experiments. The utility of the software was tested by conducting comparisons between semi-targeted and untargeted data-dependent analysis experiments on a variety of proteins, including IgG, a protein whose glycosylation must be characterized during its production as a biotherapeutic. When the GlycoPep MassList software was used to generate inclusion lists for LC-MS/MS experiments, more unique glycopeptides were selected for fragmentation. Generally, ~30 % more unique glycopeptides can be analyzed per protein, in the simplest cases, with low background. In cases where background ions from proteins or other interferents are high, usage of an inclusion list is even more advantageous. The software is free and publicly accessible.

In another research project, we describe a unique flow cytometer (TDI SFC) that combines the high spectral resolution of spectral flow cytometry (SFC) with a CCD operated in time-delayed integration (TDI) mode for the automated immunophenotyping of rare, low abundant cells. A microfluidic device providing 1-D focusing was used to sheath cells through a 488 nm laser excitation beam. Using epi-illumination, a spectrograph was included into the emission optical path to spectrally disperse the emission along one axis of a CCD camera. The parallel shift rate of the CCD was synchronized to the cell travel through the field-of-view, which was defined by the excitation volume. This TDI SFC format allowed the CCD shutter to remain open during signal acquisition and as such, the duty cycle was ~100% allowing for rare cells to not be missed. Fluorescent calibration beads were used to optimize synchronization of the CCD's TDI clocking with the sheathed cell velocity, TDI SFC sensitivity, excitation power intensity, epi-illumination objective's numerical aperture, and total integration time. TDI integrated signals of 10^6 counts at a signal-to-noise ratio (SNR) of 610 for beads corresponding to a load of 4×10^5 antibodies per bead was achieved. Additionally, we evaluated the multiplexing capabilities by performing spectral deconvolution. Finally, a proof-of-concept application was undertaken to immunophenotype rare cells, specifically leukemic cells circulating in the blood of patients with B-cell acute lymphoblastic leukemia (B-ALL) for monitoring measurable residual disease (MRD). A B-ALL cell line was stained against a leukemic marker (TdT) to successfully discriminate TdT(+) circulating leukemic cells from normal B cells at very low cell counts ($\leqslant$ 100 cells).

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my research supervisor Dr. Soper for his generous support, valuable guidance, and tremendous encouragement. His enthusiasm and knowledge in science and engineering will forever inspire me in my research career. I also want to sincerely thank my supervisor Dr. Desaire for her immense effort and patience when introducing me to the world of bioanalytical chemistry. Her generous support not only in research but also in my personal life of being a new mother serves as an invaluable part of my life in graduate school at KU. I especially want to thank Dr. Matt Jackson for his guidance and for closely working with me. Dr. Jackson's positive attitude, broad knowledge, and hands-on experience not only encourages everyone in the group, but also has impacted me in many positive ways. I also want to thank Dr. Malgorzata Witek, Dr. Nicholas Larky, Dr. Kumuditha Ratnayake, Dr. Camila Campos from Dr. Soper group, Dr. Mengjia Hu from Biofluidica, Dr. Eden Go, Dr. Zhikai Zhu, and Dr. Jude Lakbub from Dr. Desaire group for their valuable research advice and in-lab training. My previous colleagues, including Dr. Kasun Imaduwage, Joshua Shipman and Milani Wijeweera Patabandige, not only provided a lot of thought-provoking ideas for my project, but also offered spiritual support to me when my research was not going as I expected. I sincerely thank them for their help and kindness.

I especially want to thank my beloved husband Dr. Russell Fulmer, who gave me tremendous support in various ways. As a wonderful husband, a great father, and an excellent psychologist, he takes care of me and our son Luke. He is always willing to listen to me and gives me strength when things are going downhill. I want to thank my son Luke for giving me the precious opportunity of being a mother. I'm always amazed by how much strength this little guy can give me. I also want to thank my extended family members such as my parents-in-law

Russ and Roxey Fulmer and my American grandma Phyllis Fulmer for their selfless help and support through the years.

My gratitude to my family in China is beyond words. I want to thank my father Songbai Hu, my mother Xiaoping Zhang, my younger brother Weijun Hu, my uncle Bailin Hu and aunt Fen Zhang, as well as my cousins for their endless love and support.

A lot of thanks go to my friends, including Xiaomeng, Lin, Yunan, Yan, Yuqi and Zheng for all the nice time we spent together. I still remember the happy and fun talks I had with an old friend Andrea although he is now far away in Italy. Thank you, friends.

And thank you, my Lord, for your gracious love!

Table of Contents

List of Figures

List of Tables

CHAPTER I

Introduction to Analysis of Glycoproteins

1.1 Protein Glycosylation

Carbohydrates, lipids, proteins, and nucleic acids are the four major families of macromolecules in cells. Among them, proteins exhibit the most diversity in structure and function and participate in almost every biochemical process in cells.[1] One reason for the multiple and variant roles proteins are playing is post translational modifications (PTMs). PTMs are covalent modification processes that occur in polypeptides, which often requires relevant enzymes.[2]

One important and diverse PTM is glycosylation, which occurs in more than half of eukaryotic proteins.[3-5] Glycosylation refers to the process in which oligosaccharide moieties are covalently attached to proteins. The oligosaccharides, also named as glycans, play significant roles in protein structure stabilization and protein functions. For example, it is reported that the hydrophilic hydroxyl group-rich glycans enhance the solubility and stability of proteins and protect them from enzymatic degradation.[6] Many proteins, especially cell surface proteins, are glycosylated because glycoproteins as informational entities, play crucial role in cell-matrix signaling [7,8], cell-cell interactions and immune response [9,10].

Beside the important roles of naturally produced glycoproteins in eukaryotes, the significance of recombinantly expressed glycoprotein therapeutics is also tremendous. As protein therapeutics, more and more recombinant monoclonal antibodies (rMAbs) from the immunoglobulin G (IgG) family are licensed and produced due to their clinical significance and the considerable revenue brought to the pharmaceutical industry.[11] An ideal monoclonal antibody therapeutic should have ideal structural stability, good efficacy for a specific disease, and limited

side effects. Almost all of these features heavily depend on the monoclonal antibody drug's glycosylation profile.[11,12] For example, antibodies with high mannose glycans are cleared faster in human serum.[13] Glycans containing NeuGc can cause an unwanted immune response.[12] Therefore, characterization of glycoproteins has attracted much attention due to their great potential in unraveling the mechanisms of many biological processes and developing new pharmaceuticals.[7,8,14-16]

The building blocks of glycans include but are not limited to mannose, galactose, fucose, N- acetylglucosamine and N-acetylneuraminic acid. In order to present the sugar residues in an efficient way, a set of colorful symbols proposed by the Consortium for Functional Glycomics to represent monosaccharide residues are applied to simplify the structural representations of glycans.[17] Some of these symbols and the monosaccharide residues they represent as well as their abbreviations are shown in **Figure 1.1**.

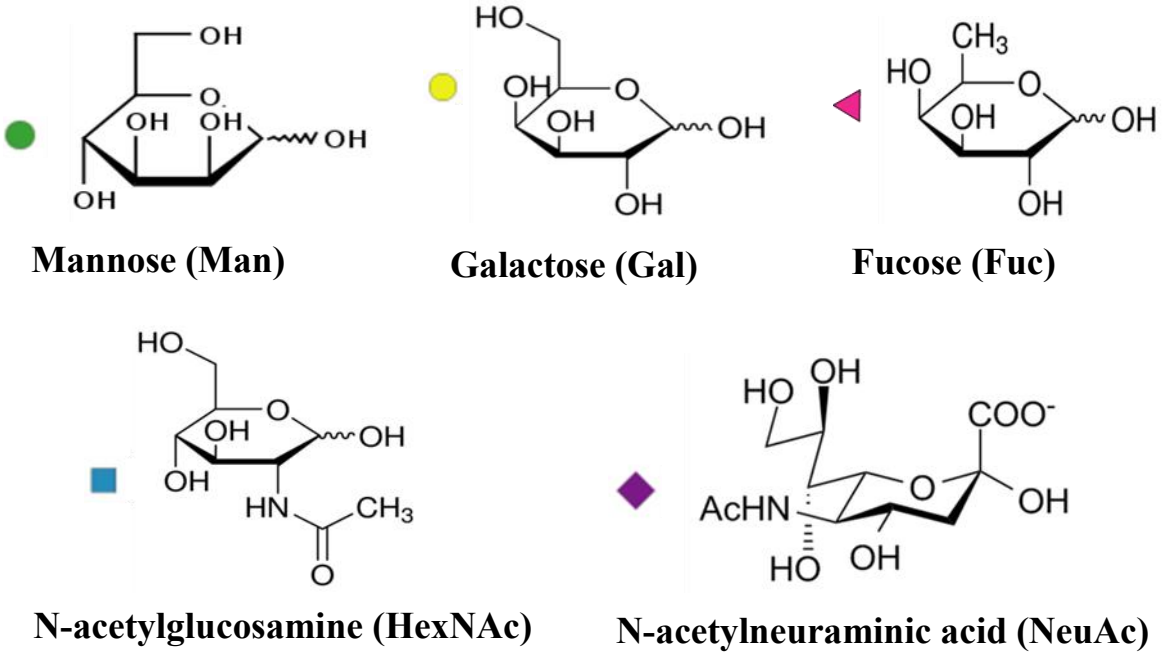


Figure 1.1 Monosaccharide residues' chemical structures, symbols, names and abbreviations.

1.1.1 N-linked Glycosylation

In this thesis, only N-linked glycosylation is studied. N-linked glycosylation is one of the most investigated glycosylation types among the existing forms of glycosylation in nature.[18] In N-linked glycosylation, the glycan forms a covalent bond with the side-chain amine group of an asparagine residue (Asn or N) following a consensus amino acid sequence of NXS/T (where X can be any amino acid residue except proline).[19] It should be noted that although all N-linked glycoproteins require such a preserved amino acid sequence, the presence of such peptide backbone sequence does not necessarily guarantee glycosylation at that site.[20] This phenomenon also extends to other types of glycosylation.

The process of glycosylation depends on the protein's microenvironment and enzyme availability as opposed to peptide backbone biosynthesis controlled by the genome. Each mature N-linked glycoprotein stems from a precursor glycoprotein linked with a precursor glycan of $Glc_3Man_9GlcNAc_2$. The precursor glycoprotein is biosynthesized in the endoplasmic reticulum (ER) lumen under the catalysis of multiple oligosaccharide transferases (OSTase).[21] Mature glycans are formed by truncating and extension of the precursor glycan in ER and/or Golgi apparatus.[21] Despite the fact that trimming and extension happen to the precursor glycoprotein, the core moiety of five oligosaccharide residues linked to the asparagine residue is usually retained. This gives rise to the preserved core motif of two N-acetyl glucosamine residues and three mannose residues in N-linked glycans[22], as shown in **Figure 1.2**.

Depending on the compositions of the N-glycans, they can be divided into three subgroups: high mannose, complex, and hybrid as shown in **Figure 1.2**. A high mannose glycan is formed by removing the mannoses from the precursor glycan without adding new monosaccharides. A complex glycan is formed by removing and adding sugars to the precursor

glycan. A hybrid glycan, as self-explained by its name, is a combination of the building blocks of both high mannose and complex glycans.[22]

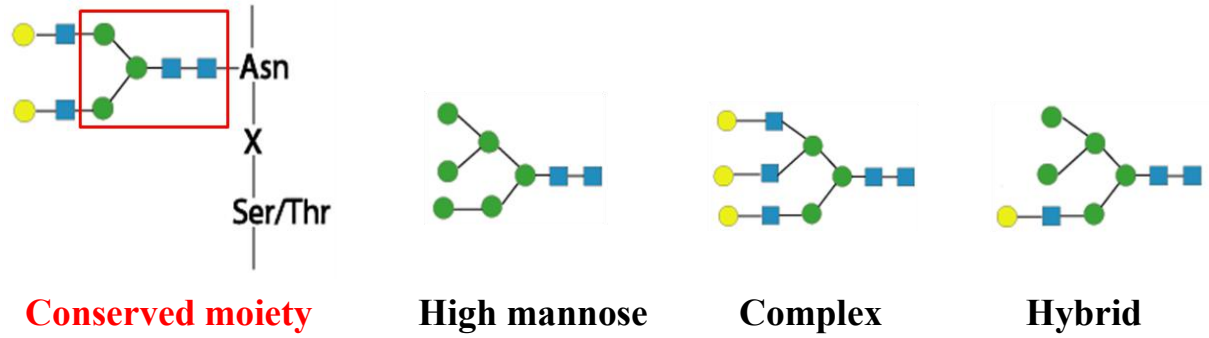


Figure 1.2 N-linked glycans have a conserved moiety of $Hex_3HexNAc_2$. All N-linked glycans can be divided into three subgroups, high mannose, complex and hybrid.

1.1.2 Other Types of Glycosylation

Besides N-linked glycosylation, there are other forms of glycosylation. Another widely studied type is O-linked glycosylation. In O-linked glycosylation, glycans are attached to the hydroxyl group of a serine or threonine residue (preserved amino acid sequence S/T).[23] Less common glycosylations include glypiation[24], C-glycosylation[25], and phosphoglycosylation[26].

1.1.3 Heterogeneity of Glycosylation

Unlike the biosynthesis of a protein peptide backbone, which is templated by the genome, the process of glycosylation is not a templated process. Instead, it is largely determined by the availability of related enzymes, such as glycosyltransferase and glycosidase. This results in heterogeneity of the glycoprotein population.[22,27] An example of the macroheterogeneity and the microheterogeneity of a glycoprotein is illustrated in **Figure 1.3**.

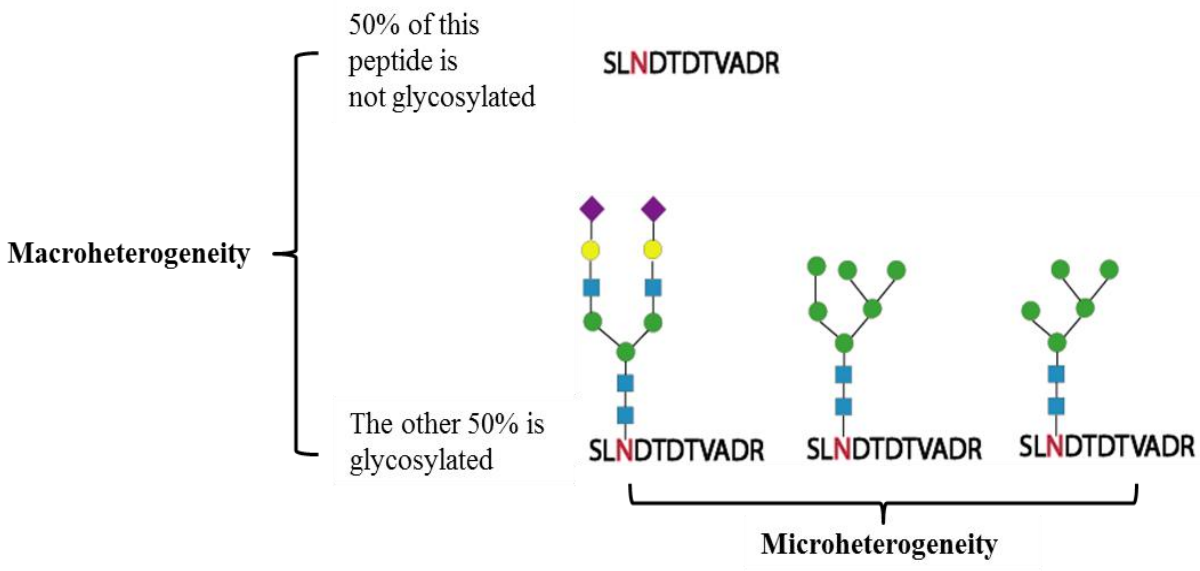


Figure 1.3 Macroheterogeneity and microheterogeneity of glycosylation

This variance of glycosylation site occupancy is called macroheterogeneity.[28] With macroheterogeneity, for one specific glycosylation site, the degree of glycosylation may not be 100%. The site may or may not be glycosylated, depending on the different copies of the proteins. Microheterogeneity, on the other hand, corresponds to the attachment of different glycans to the same glycosylation sites in different copies of proteins.[28] Heterogeneity renders glycoproteins in low abundance and thus makes glycoprotein/glycopeptide analysis challenging.

1.1.4 Enrichment of Glycoproteins/glycopeptides

In order to alleviate the difficulty brought about by the low abundance of glycoproteins, enrichment strategies for glycoproteins or glycopeptides are frequently utilized. The positive selection and enrichment greatly reduces the sample complexity, decreasing the interference brought about by non-glycosylated substrates.[29]

Enrichment can be categorized into molecular interaction and chemical reaction based methods.[29] Molecular interactions include hydrophilic interaction, lectin affinity interaction, and titanium dioxide-based coordination interaction while chemical reactions encompass hydrazide

and boronic acid chemistry.[29] Among the former group of enrichment methodologies/methods, hydrophilic chromatography and lectin affinity chromatography are frequently employed for N-linked glycopeptide enrichment.[30] In hydrophilic interaction liquid chromatography (HILIC), multiple hydroxyl groups-containing glycopeptides partition between the hydrophobic mobile phase and hydrophilic stationary phase, allowing for enrichment.[31] However, the non-specific hydrophilic interaction can result in co-enrichment of many polar groups containing compounds from the sample matrix, which dampens the goal of glycopeptide enrichment.[30] Lectins are proteins that have selective recognition and binding for certain groups of oligosaccharide moieties.[32] The commonly used lectins are concanavalin A (Con A) mainly with specificity for mannose, jacalin for galactosyl (β1-3) N-acetylgalactosamine and wheat germ agglutinin (WGA) for N-acetyl glucosamine and sialic acid.[32] Although lectin affinity chromatography (LAC) has the feature of high specificity, the preferential affinity to a subset of glycans can cause sample loss.[33] Chemical reaction based glycopeptide enrichment has less discrimination against the subgroups of glycans.[34] However, the recovery is largely dependent on the releasing step.[32] In addition, hydrazide chemistry detached the glycan parts from the peptides, hindering the localization of glycosites.[34]

In summary, although glycopeptide analysis benefits from glycopeptide enrichment, the weak binding, biased affinity and multiple sample preparation steps hamper the purpose of nondiscriminatory identification and impede the goal for high glycopeptide coverage.

1.2 Analysis of Glycoproteins

Since glycoproteins are involved with proteins and glycans, the analysis of glycoproteins is more challenging than the analysis of mere proteins or glycans. In addition, heterogeneity of glycoproteins adds one more dimension of complexity and difficulty to the analysis. Generally,

the analysis of glycoproteins can be achieved by non-mass spectrometry based approaches and mass spectrometry based approaches.

1.2.1 Non-mass spectrometry Approaches

The most common non-MS based approaches include but are not restricted to sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) staining and affinity selection coupled with fluorescence detection.[23] These approaches tell whether a protein has been glycosylated.

Gel staining with SDS-PAGE uses periodic acid to oxidize two vicinal diol groups to form an aldehyde group, which reacts with a Schiff reagent to produce a magenta color.[35] In another strategy, enzyme Peptide-N-glycosidase F (PNGase F) is used to deglycosylate the protein, which shows mass differences with the PNGase F-untreated protein on SDS-PAGE, to reveal if a protein is glycosylated.[36]

In affinity based procedures, glycoproteins are captured by immobilized antibodies, which is followed by the addition of an enzyme conjugated secondary antibody. The affinity selection coupled with fluorescence generation catalyzed by the enzyme enables the determination of glycosylation.[23]

Other non-MS methods include HPLC separation coupled with fluorescence detection.[37,38] Most naturally produced glycans possess minimal to no fluorescence. But they can be derivatized with fluorescent labels, such as 2-amino benzamide (2 AB) or 2-anthranilic acid (2 AA), and make fluorescence detection feasible.[38]

The non-MS based methods are relatively simple and straightforward. However, they lack the ability of giving information on the peptide backbone sequence or the glycan compositions. To obtain this valuable information, mass spectrometry is often utilized.

1.2.2 Mass Spectrometry Based strategies

Over the past decades, mass spectrometry (MS) has demonstrated its invaluable merits in characterizing PTMs[39-41], including glycosylation[42,43]. A frequently adopted workflow for glycoprotein analysis using MS involves bottom-up proteomics. In bottom-up proteomics, peptides generated after enzymatic digestion are subjected to High Performance Liquid Chromatography (HPLC) tandem mass spectrometry (MS/MS) for in-depth analysis.[44,45]

Regardless of the type of mass spectrometer being used, a mass spectrometer usually consists of the following three major units: (1) an ion source to ionize the analytes, (2) a mass analyzer to separate the analytes based on different mass-to-charge ratios (m/z), and (3) a mass detector to detect the analytes. In order to be analyzed by a mass spectrometer, the analytes must be ionized. Therefore, the ion source is very important because it determines what type of analytes or application is suitable for the specific mass spectrometer. There are different ion sources such as electron impact (EI), chemical ionization (CI), fast atom bombardment (FAB)[46], electrospray ionization (ESI)[47] and matrix-assisted laser desorption ionization (MALDI)[48]. The later three are known as "soft ionization" methods due to their mild pathways to ionize biological molecules, such as proteins and deoxyribonucleic acids (DNA), without remarkable decomposition of molecular structure.[49,50]

MS first successfully identified a glycopeptide primary structure in 1970s by using EI and CI.[51] The glycans from the glycopeptides were released and derivatized to increase the volatility so that they could be detected by MS.[51] In the late 1980s and early 1990s, the sprout and development of "soft ionization" techniques, electrospray ionization (ESI)[47,52] and matrix-assisted laser desorption ionization (MALDI)[48,53], promoted mass spectrometry into a revolutionary area where analysis of biological macromolecules by MS became much more

feasible[54,55]. John Fenn and Koichi Tanaka, pioneers of ESI and MALDI, respectively, were awarded the 2002 Noble Prize in Chemistry for their significant contribution to the field. In ESI and MALDI, the non-volatile glycoproteins or glycopeptides can be ionized without significant structure disruption or loss of labile groups.[56,57] Because of the commercial availability of ESI and MALDI, these two soft ion sources have been frequently applied to analyze glycoconjugates and glycoproteins.[9,25,58]

Particularly, the marriage of ESI and HPLC further boosted mass spectrometry's capability for the analysis of glycoproteins.[59-62] The online separation provided by HPLC coupled with mass detection increases the data and information obtained in one single experiment and enables the analysis of complex biological samples.[63,64]

1.2.2.1 Glycopeptide Analysis by MS1

After the glycopeptides are ionized, they can be separated by mass analyzers such as a quadrupole ion trap[65], orbitrap[66], Fourier Transform ion cyclotron resonance (FTICR)[67], or Time-of Flight (TOF) spectrometer. In mass analyzers engineered with electric and magnetic fields, ions with different m/z values deflect at different degrees allowing separation of them.[50] Due to the different features of these mass analyzers and the detectors coupled to them, they can be quite different in terms of cost, speed, resolution and dynamic range. For example, the quadrupole ion trap is best known for its fast scan rate whereas orbitrap is frequently used for its high resolution data.

High resolution ESI MS1 data is very useful in glycopeptide analysis as the mass error within several ppm can help to define the possible glycopeptide compositions.[68] Nevertheless, high mass accuracy data, in some cases, are not sufficient to rigorously identify the glycopeptide because there are numerous combinations of amino acids with oligosaccharides. The isobaric

MS1 data may result in misclassification of glycopeptides and lead to wrong conclusions.[68] In this case, MS2 or MS/MS data are needed to verify the analysis and eliminate the possibility of misidentification.

1.2.2.2 Glycopeptide Analysis by MS/MS

Collision induced dissociation (CID) is a widely used dissociation method, where glycosidic fragment ions as well as peptide backbone b and y ions are produced due to vibrational activation[69], enabling the identification of glycopeptides. In this low-energy pathway dominated process, the predominant dissociation of glycan parts gives rise to oxonium ions (NeuAc+, m/z 292; HexHexNAc+, m/z 366; HexHexNAcNeuAc+, m/z 657)[70] that are diagnostic ions of glycopeptides[19]. Another advantage of CID is the fast scan speed enabled by the ion trap[65], allowing for simultaneous elucidation of multiple glycopeptides sharing the same peptide sequence but different glycans when coeluting in a reverse phase LC column[71]. However, despite the ability to precisely identify glycopeptides, glycopeptide analysis by CID still suffers from the intrinsic low abundance nature of glycopeptides. The macroheterogeneity that comes from the difference of glycosylation site occupancy and the microheterogeneity that results from the attachment of different glycans to one specific glycosylation site render each glycoform in a low copy number after proteolysis.[5,28] Consequently, in the most commonly used MS/MS data dependent acquisition (DDA) mode where the most intense peptide ions on MS are selected for MS/MS, the high-abundance precursor ions are redundantly selected for MS/MS while the relatively low abundant glycopeptide ions may not trigger MS/MS even if dynamic exclusion is enabled.[72] This results in missed detection of glycopeptides and limited glycopeptide coverage. This hurdle is especially serious when the sample has a complex matrix that brings high background corresponding to non-glycosylated peptides. The goal of the research work discussed

in Chapter 3 is to alleviate this problem by developing a new software tool and a different MS/MS strategy from the most commonly used one.

Electron transfer dissociation (ETD) is another commonly used fragmentation method for glycopeptide analysis. Unlike CID, ETD predominantly generates fragment ions from peptide back bone cleavage.[73] This is due to its radical induced fragmentation pathways instead of vibrational energy controlled pathways in CID. **Figure 1.4** exhibits the difference of fragmentation behavior of CID and ETD.

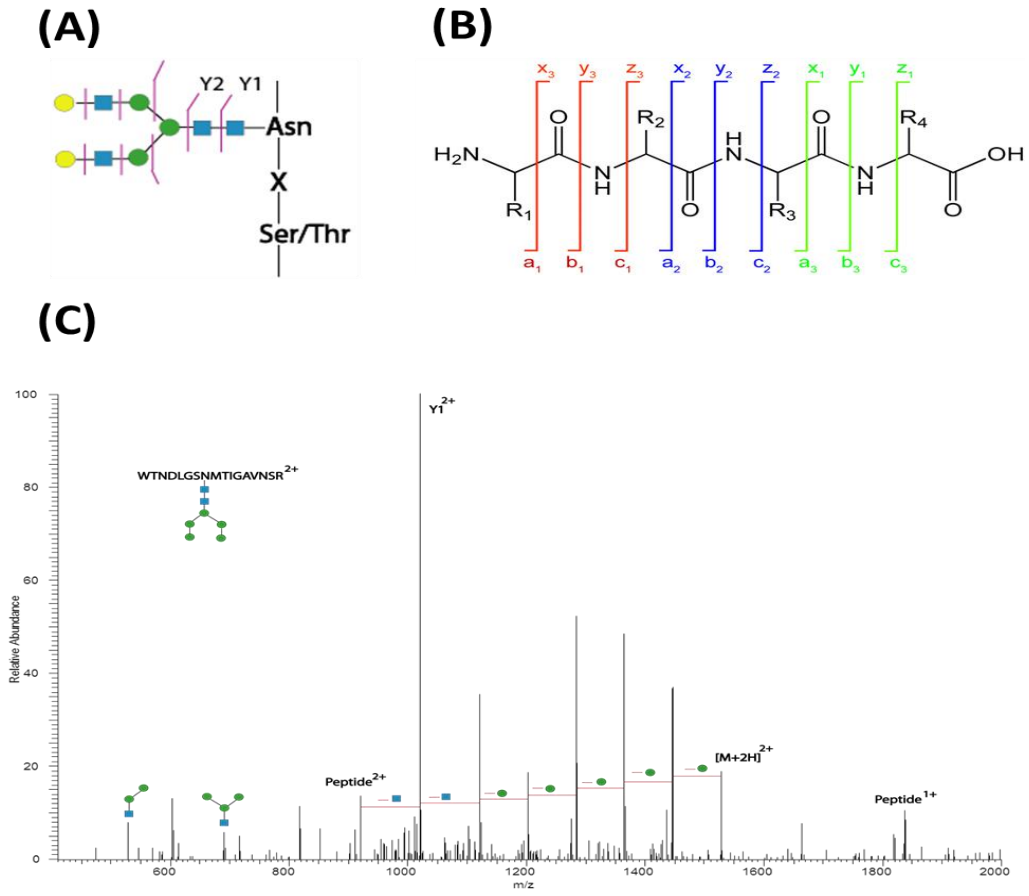


Figure 1.4 (A) In CID fragmentation mainly occurs at glycosidic bonds. (B) In ETD dissociation primarily occurs at peptide backbone bonds. (C) CID MS/MS data of an N-linked high-mannose avidin glycopeptide (z=2, m/z=1527.66).

Besides CID, new MS/MS strategies have also been developed for glycopeptide identification. Hart-Smith and colleagues investigated using high resolution MS1 and higher

energy C-trap dissociation (HCD) to identify N-linked glycopeptides from a hen egg glycoprotein mixture.[74] The HCD MS/MS detected by orbitrap was featured with high resolution data and the absence of 1/3 low-mass cutoff found in the ion trap. The high mass accuracy of the Y1 ions (peptide + HexNAc ions) and oxonium ions from HCD data ensured the confirmation of glycopeptides.[74]

Singh and coworkers performed ETD triggered by HCD oxonium product ions to analyze glycopeptides.[75] In their approach, when the oxonium peaks (HexNAc+, m/z 204.09; HexHexNAc+, 366.14) were among the top 20 intense fragment peaks on HCD MS/MS, a subsequent ion trap ETD scan would be performed on the same precursor ion.[75] HCD mainly generated the glycosidic fragment peaks.[76] ETD primarily produced peptide backbone c and z ions with the heterogeneous glycans largely retained, enabling peptide backbone sequencing and glycosite localization.[77] The complementary information obtained from HCD and ETD served a robust tool for glycopeptide analysis.[75] However, the high resolution orbitrap HCD MS/MS scans had longer duty cycle in comparison with ion trap CID MS/MS, resulting in less HCD MS/MS scans during the same time. In addition, in the abovementioned strategies involving HCD the commonly used DDA was used, which meant only the top intense precursor ions on MS1 preview scan were selected for MS/MS. Nevertheless, because of the low abundance of glycopeptides most of the selected precursor ions for MS/MS in DDA mode did not correspond to glycopeptides. The long cycle time and biased ion selection can be detrimental if broad scale glycopeptide identification is desired.

1.3 Challenges in MS/MS Analysis of Glycopeptides

During the commonly used MS/MS data collection process where the DDA mode is used, the relatively low abundance glycopeptide ions may not trigger MS/MS in this intensity based

ion selection process. This results in the absence of MS/MS data of some glycopeptides, leading to limited glycopeptide coverage, as indicated by **Figure 1.5**.

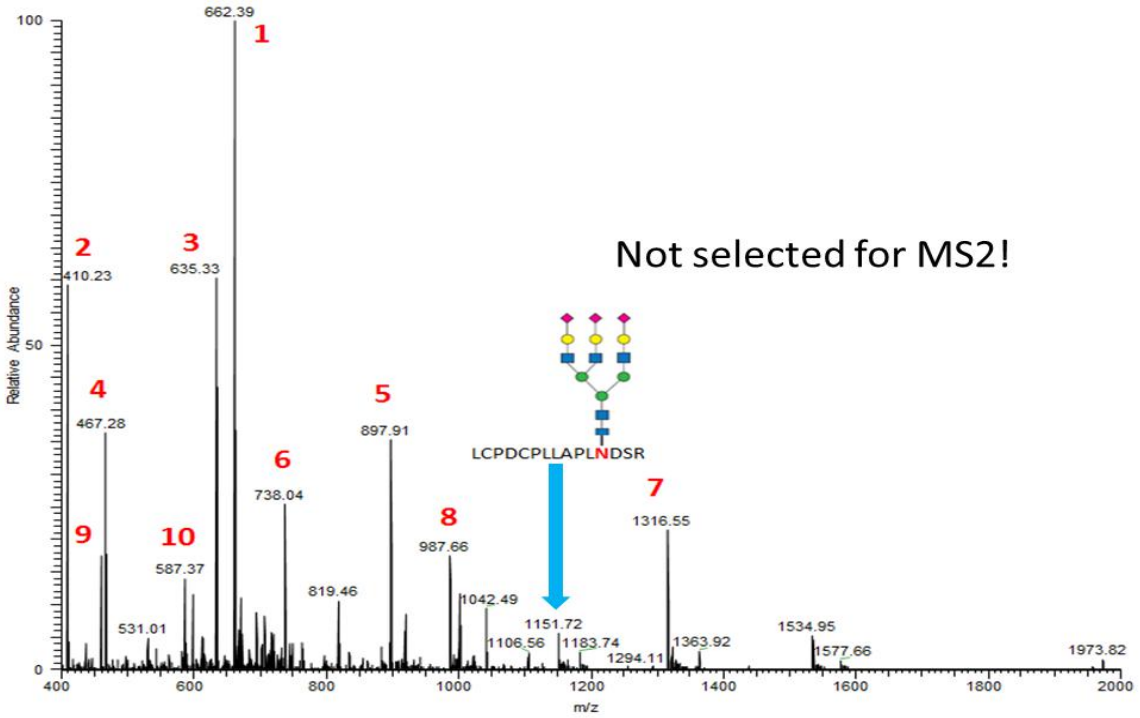


Figure 1.5 In the commonly used DDA mode the most abundant parent ions, for instance, the top 10 ions indicated by red numbers, on a MS1 scan are selected for MS/MS. The glycopeptide ions with low abundance, such as the one indicated with blue arrow, are not selected for MS/MS scan.

1.4 Conclusion

Analyzing glycopeptides by MS1 is essential. Due to the numerous combination of glycans and peptide backbones, analyzing glycopeptide merely by MS1 data is not adequate to confirm the analytes due to the numerous possible isobaric glycopeptides. Hence, MS/MS data are usually needed to accurately verify the analysis. In the commonly used DDA mode during MS/MS data acquisition, the ion selection for MS/MS is based on intensity of the parent peak on MS1 scan. When there are low-abundance glycopeptide species, they may not trigger MS/MS due to their low peak intensity on MS1, decreasing the glycopeptide coverage. Semi targeted data acquisition is an alternative that can alleviate this issue. In this approach, a list of pre-

assigned glycopeptide masses is used to trigger MS/MS. However, accurately generating a large number of glycopeptide masses can be time-consuming. Hence, new software tools is needed for facilitating this strategy to circumvent the challenges in DDA.

1.5 Protein Analysis vs Cell Analysis

Analysis of glycoprotein by MS has attracted lots of researcher interest because glycoproteins are not only involved in many biological processes, such as fertilization[78], inter-cell signaling[32], and antibody-antigen bindings[79], but also serve as potential disease markers [80,81]. For example, increased core fucosylation of alpha-fetoglycoprotein (AFP) was reported to associate with hepatocellular carcinoma (HCC), a type of liver cancer, and thus could be used as a biomarker for HCC in patients infected with Hepatitis B virus (HBV) or/and Hepatitis C virus (HCV).[81]

Although glycoproteins may serve as potential disease detection markers, what is needed is a biomarker to track disease progression. Many times the disease detection marker may not necessarily enable the knowledge of disease progression. For instance, increased core fucosylation on alpha-fetoglycoprotein (AFP) indicates highly possible HCC in HBV or/and HCV infected patients. However, the information provided by MS is insufficient to provide knowledge about the progression or severity of the disease. Directly analyzing the anatomically accessible specimens is capable of unraveling more information related to the number and/or percentage of the malignant cells and enables the knowledge of the disease progression or severity. Due to its unique physical features, MS is unable to analyze anatomically accessible specimens. In this situation, other bioanalytical methods that are able to analyze tissue or cell samples from cancer patients are highly needed. Flow cytometry is one of such bioanalytical methods that can analyze cells from patients, of which the majority are leukemia patients.

1.6 References

(1) Dwek, R. A. *Chemical Reviews* **1996**, *96*, 683-720.

(2) Santos, A. L.; Lindner, A. B. *Oxidative Medicine and Cellular Longevity* **2017**, *2017*, 1-19.

(3) Spiro, R. G. *Glycobiology* **2002**, *12*, 43-56.

(4) Ohtsubo, K.; Marth, J. D. *Cell* **2006**, *126*, 855-867.

(5) Aldredge, D.; An, H. J.; Tang, N.; Waddell, K.; Lebrilla, C. B. *Journal of proteome research* **2012**, *11*, 1958-1968.

(6) Sola, R. J.; Griebenow, K. *Journal of Pharmaceutical Sciences* **2009**, *98*, 1223-1245.

(7) Ohtsubo, K.; Marth, J. D. *Cell* **2006**, *126*, 855-867.

(8) Gaucher, S. P.; Leary, J. A. *J Am Soc Mass Spectrom* **1999**, *10*, 269-272.

(9) Royle, L.; Roos, A.; Harvey, D. J.; Wormald, M. R.; van Gijlswijk-Janssen, D.; Redwan el, R. M.; Wilson, I. A.; Daha, M. R.; Dwek, R. A.; Rudd, P. M. *The Journal of biological chemistry* **2003**, *278*, 20140-20153.

(10) Talbot, P. *Biology of Reproduction* **2002**, *68*, 1-9.

(11) Jefferis, R. *Nature reviews. Drug discovery* **2009**, *8*, 226-234.

(12) Ghaderi, D.; Taylor, R. E.; Padler-Karavani, V.; Diaz, S.; Varki, A. *Nature biotechnology* **2010**, *28*, 863-867.

(13) Shah, B.; Jiang, X. G.; Chen, L.; Zhang, Z. *J. Am. Soc. Mass Spectrom.* **2014**, *25*, 999-1011.

(14) Foudaa, G. G.; Jaegera, F. H. *PNAS* **2013**, *110*, 18220-18225.

(15) Hong, P.; Ninonuevo, M. R.; Lee, B.; Lebrilla, C.; Bode, L. *British Journal of Nutrition* **2009**, *101*, 482-486.

(16) Drake, P. M.; Schilling, B.; Niles, R. K.; Prakobphol, A.; Li, B.; Jung, K.; Cho, W.; Braten, M.; Inerowicz, H. D.; Williams, K.; Albertolle, M.; Held, J. M.; Iacovides, D.; Sorensen, D. J.;

Griffith, O. L.; Johansen, E.; Zawadzka, A. M.; Cusack, M. P.; Allen, S.; Gormley, M., et al. *Journal of proteome research* **2012**, *11*, 2508-2520.

(17) Raman, R.; Venkataraman, M.; Ramakrishnan, S.; Lang, W.; Raguram, S.; Sasisekharan, R. *Glycobiology* **2006**, *16*, 82-90.

(18) Morelle, W.; Canis, K.; Chirat, F.; Faid, V.; Michalski, J. C. *Proteomics* **2006**, *6*, 3993-4015.

(19) Mechref, Y. *Current protocols in protein science* **2012**, *12*, 1-11.

(20) Nita-Lazar, M.; Wacker, M.; Schegg, B.; Amber, S.; Aebi, M. *Glycobiology* **2005**, *15*, 361-367.

(21) Nakayama, Y.; Nakamura, N.; Tsuji, D.; Itoh, K.; Kurosaka, A. In *Genetic Disorders*, Puiu, M., Ed.; IntechOpen, 2013, pp 243-269.

(22) Dalpathado, D. S.; Desaire, H. *Analyst* **2008**, *133*, 731-738.

(23) Roth, Z.; Yehezkel, G.; Khalaila, I. *International Journal of Carbohydrate Chemistry* **2012**, *2012*, 1-10.

(24) Walker-Nasir, E.; Ahmad, I.; Saleem, M.; Hoessli, D. C. *Current Organic Chemistry* **2007**, *11*, 591-607.

(25) Hofsteenge, J.; Muller, D. R.; Debeer, T.; Loffler, A.; Richter, W. J.; Vliegenthart, J. F. G. *Biochemistry* **1994**, *33*, 13524-13530.

(26) Haynes, P. A. *Glycobiology* **1998**, *8*, 1-5.

(27) Go, E. P.; Rebecchi, K. R.; Dalpathado, D. S.; Bandu, M. L.; Zhang, Y.; Desaire, H. *Anal Chem* **2007**, *79*, 1708-1713.

(28) Li, F.; Glinskii, O. V.; Glinsky, V. V. *Proteomics* **2013**, *13*, 341-354.

(29) Chen, C. C.; Su, W. C.; Huang, B. Y.; Chen, Y. J.; Tai, H. C.; Obena, R. P. *Analyst* **2014**, *139*, 688-704.

(30) Yu, L.; Li, X.; Guo, Z.; Zhang, X.; Liang, X. *Chemistry* **2009**, *15*, 12618-12626.

(31) Buszewski, B.; Noga, S. *Anal Bioanal Chem* **2012**, *402*, 231-247.

(32) Frost, D. C.; Li, L. *Adv. Protein Chem. Struct. Biol.* **2014**, *95*, 71-123.

(33) Zhang, Y.; Jiao, J.; Yang, P.; Lu, H. *Clinical proteomics* **2014**, *11*, 1-14.

(34) Ongay, S.; Boichenko, A.; Govorukhina, N.; Bischoff, R. *Journal of Separation Science* **2012**, *35*, 2341-2372.

(35) Zachariu.Rm; Zell, T. E.; Morrison, J. H.; Woodlock, J. J. *Analytical biochemistry* **1969**, *30*, 148-152.

(36) Wilson, N.; Simpson, R.; Cooper-Liddell, C. *Methods in Molecular Biology, Glycomics: Methods and Protocols* **2009**, *534*, 205-212.

(37) Kozak, R. P.; Tortosa, C. B.; Fernandes, D. L.; Spencer, D. I. *Analytical biochemistry* **2015**, *486*, 38-40.

(38) Bigge, J.; Patel, T.; Bruce, J.; Goulding, P.; Charles, S.; Parekh, R. *Analytical biochemistry* **1995**, *230*, 229-238.

(39) Olsen, J. V.; Macek, B.; Lange, O.; Makarov, A.; Horning, S.; Mann, M. *Nat Methods* **2007**, *4*, 709-712.

(40) Zhao, Y.; Jensen, O. N. *Proteomics* **2009**, *9*, 4632-4641.

(41) Drake, S. K.; Hortin, G. L. *The international journal of biochemistry & cell biology* **2010**, *42*, 174-179.

(42) Udiavar, S.; Apffel, A.; Chakel, J.; Swedberg, S.; Hancock, W. S.; Pungor, E. *Analytical Chemistry* **1998**, *70*, 3572-3578.

(43) Wang, H.; Wong, C.-H.; Chin, A.; Taguchi, A.; Taylor, A.; Hanash, S.; Sekiya, S.; Takahashi, H.; Murase, M.; Kajihara, S.; Iwamoto, S.; Tanaka, K. *Nature Protocols* **2011**, *6*, 253-269.

(44) Peterman, S. M.; Mulholland, J. J. *J Am Soc Mass Spectrom* **2006**, *17*, 168-179.

(45) Xu, P.; Duong, D. M.; Peng, J. *Journal of proteome research* **2009**, *8*, 3944-3950.

(46) Dell, A.; Morris, H. R.; Easton, R. L.; Panico, M.; Patankar, M.; Oehninger, S.; Koistinen, R.; Koistinen, H.; Seppala, M.; Clark, G. F. *Journal of Biological Chemistry* **1995**, *270*, 24116-24126.

(47) Fenn, J. B.; Mann, M.; Meng, C. K.; Wong, S. F.; Whitehouse, C. M. *Science* **1989**, *246*, 64-71.

(48) Tanaka, K.; Waki, H.; Ido, Y.; Akita, S.; Yoshida, Y.; Yoshida, T.; Matsuo, T. *Rapid Communications in Mass Spectrometry* **1988**, *2*, 151-153.

(49) Sasaki, H.; Ochi, N.; Dell, A.; Fukuda, M. *Biochemistry* **1988**, *27*, 8618-8626.

(50) El-Aneed, A.; Cohen, A.; Banoub, J. *Applied Spectroscopy Reviews* **2009**, *44*, 210-230.

(51) Morris, H. R.; Thompson, M. R.; Osuga, D. T.; Ahmed, A. I.; Chan, S. M.; Vandenheede, J. R.; Feeney, R. E. *Journal of Biological Chemistry* **1978**, *253*, 5155-5162.

(52) Smith, R. D.; Loo, J. A.; Edmonds, C. G.; Barinaga, C. J.; Udseth, H. R. *Analytical Chemistry* **1990**, *62*, 882-899.

(53) Hillenkamp, F.; Karas, M. *Methods in Enzymology* **1990**, *193*, 280-295.

(54) Han, X. L.; Gross, R. W. *Proceedings of the National Academy of Sciences of the United States of America* **1994**, *91*, 10635-10639.

(55) Loo, J. A. *Mass spectrometry reviews* **1997**, *16*, 1-23.

(56) Griffiths, J. *Anal Chem* **2008**, *80*, 5678-5683.

(57) Rahbek-Nielsen, H.; Roepstorff, P.; Reischl, H.; Wozny, M.; Koll, H.; Haselbeck, A. *Journal of Mass Spectrometry* **1997**, *32*, 948-958.

(58) Itoh, S.; Kawasaki, N.; Harazono, A.; Hashii, N.; Matsuishi, Y.; Kawanishi, T.; Hayakawa, T. *Journal of chromatography. A* **2005**, *1094*, 105-117.

(59) Karlsson, K. E.; Novotny, M. *Analytical Chemistry* **1988**, *60*, 1662-1665.

(60) Griffin, P. R.; Coffman, J. A.; Hood, L. E.; Yates, J. R. *International Journal of Mass Spectrometry and Ion Processes* **1991**, *111*, 131-149.

(61) Rohrer, J. S.; Cooper, G. A.; Townsend, R. R. *Analytical biochemistry* **1993**, *212*, 7-16.

(62) Stadlmann, J.; Pabst, M.; Kolarich, D.; Kunert, R.; Altmann, F. *Proteomics* **2008**, *8*, 2858-2871.

(63) Martin, S. E.; Shabanowitz, J.; Hunt, D. F.; Marto, J. A. *Analytical Chemistry* **2000**, *72*, 4266-4274.

(64) Dell, A.; Morris, H. R. *Science* **2001**, *291*, 2351-2356.

(65) Schwartz, J. C.; Senko, M. W.; Syka, J. E. P. *J Am Soc Mass Spectrom* **2002**, *13*, 659-669.

(66) Zubarev, R. A.; Makarov, A. *Anal Chem* **2013**, *85*, 5288-5296.

(67) Zhou, W.; Hakansson, K. *Current proteomics* **2011**, *8*, 297-308.

(68) Desaire, H.; Hua, D. *International Journal of Mass Spectrometry* **2009**, *287*, 21-26.

(69) Quan, L.; Liu, M. *Mod Chem Appl* **2013**, *1*, 1-5.

(70) Mechref, Y.; Madera, M.; Novotny, M. V. *Mass Spectrometry of Proteins and Peptides* **2009**, *492*, 161-180.

(71) Desaire, H. *Molecular & Cellular Proteomics* **2013**, *12*, 893-901.

(72) Schmidt, A.; Gehlenborg, N.; Bodenmiller, B.; Mueller, L. N.; Campbell, D.; Mueller, M.; Aebersold, R.; Domon, B. *Molecular & Cellular Proteomics* **2008**, *7*, 2138-2150.

(73) Zhu, Z.; Hua, D.; Clark, D. F.; Go, E. P.; Desaire, H. *Anal Chem* **2013**, *85*, 5023-5032.

(74) Hart-Smith, G.; Raftery, M. J. *Journal of the American Society for Mass Spectrometry* **2012**, *23*, 124-140.

(75) Singh, C.; Zampronio, C. G.; Creese, A. J.; Cooper, H. J. *Journal of proteome research* **2012**, *11*, 4517-4525.

(76) Segu, Z. M.; Mechref, Y. *Rapid communications in mass spectrometry : RCM* **2010**, *24*, 1217-1225.

(77) Catalina, M. I.; Koeleman, C. A.; Deelder, A. M.; Wuhrer, M. *Rapid communications in mass spectrometry : RCM* **2007**, *21*, 1053-1061.

(78) Grass, J.; Pabst, M.; Chang, M.; Wozny, M.; Altmann, F. *Anal Bioanal Chem* **2011**, *400*, 2427-2438.

(79) Balzarini, J. *Nature reviews. Microbiology* **2007**, *5*, 583-597.

(80) Sturiale, L.; Barone, R.; Fiumara, A.; Perez, M.; Zaffanello, M.; Sorge, G.; Pavone, L.; Tortorelli, S.; O'Brien, J. F.; Jaeken, J.; Garozzo, D. *Glycobiology* **2005**, *15*, 1268-1276.

(81) Mehta, A.; Block, T. M. *Disease Markers* **2008**, *25*, 259-265.

CHAPTER II

Introduction to Detecting Circulating Leukemia Cells for Monitoring Relapse from Minimal Residual Disease

2.1 Leukemia and Minimal Residual Disease

2.1.1 Leukemia Introduction

Inside our bones is bone marrow (BM) that contains hematopoietic stem cells (HSCs). During hematopoiesis, HSCs differentiate into common myeloid progenitor cells and lymphoid progenitor cells both of which further differentiate into red blood cells (RBCs), white blood cells (WBCs) and platelets that comprise the formed components of blood.[1,2] However, in leukemia the HSCs or progenitor cells in the BM acquire mutations and result in abnormally high numbers of white blood cells (WBCs) that carry out minimal to no useful functions.[2] These cancerous cells usually overproduce, leading to decreased number of normal RBCs, WBCs, and platelets, causing symptoms such anemia, fatigue, and easy bruising in patients.[3] Because the malignant cells do not properly perform functions like healthy WBCs do, leukemia patients also suffer from compromised immune systems.[3]

Depending on the onset and lineage of the disease, leukemia is mainly divided into four subgroups: acute myeloid leukemia (AML), chronic myeloid leukemia (CML), acute lymphocytic leukemia (ALL) and chronic lymphocytic leukemia (CLL).[4] Leukemia is usually noticed or diagnosed with an extremely high number of WBCs from a blood test.[5] To confirm the diagnosis and determine the subtype of the disease, more tests such as morphological tests[6,7] and flow cytometric tests[8,9] are usually undertaken.

Treatments for leukemia include chemotherapy and stem cell transplantation, depending

on the disease state as well as the age and health conditions of the patient.[9,10] Once the diagnosis is confirmed, induction therapy is usually applied, especially in acute leukemia patients[11], to remove as many leukemic cells as possible . Detection of leukemic cells after treatment is essential because it is critical for monitoring disease progression and guiding treatments.[12,13]

2.1.2 Detection of Leukemic Cells

2.1.2.1 Morphology

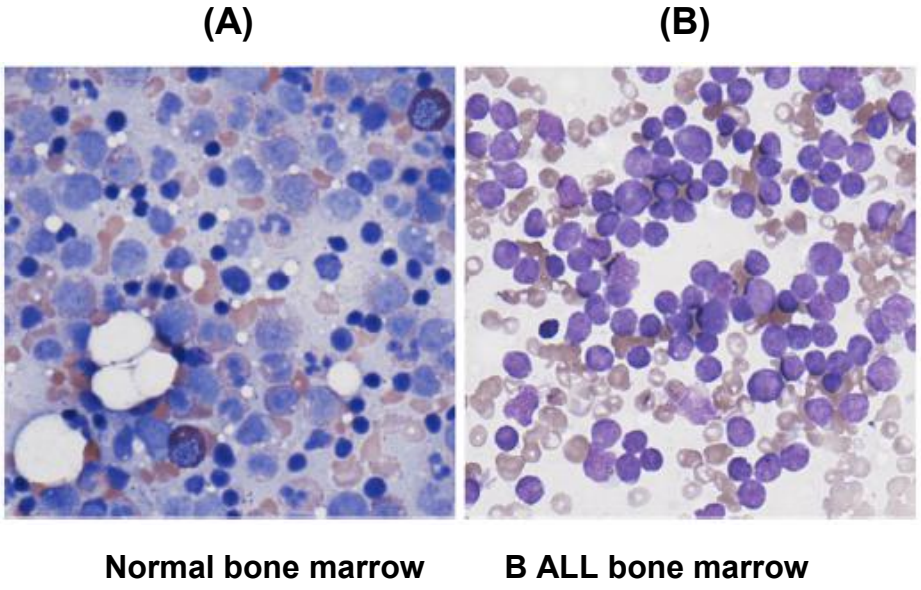


Figure 2.1 Normal bone marrow (A) contains hematopoietic cells in various stages of maturation. Bone marrow of a patient with acute leukemia (B) is filled with monotonously appearing large leukemic blasts. Reproduced from Ref. 3 with permission from The American Academy of Pediatrics.

Standard clinical tests for leukemic cells include morphology and flow cytometry.[6,14] Morphological examinations have been a requisite for leukemic cell evaluation and diagnosis[15], and post-remission disease monitoring.[16] Morphological examinations mainly provide information on the leukemic cells' size and appearance, which is evaluated using an optical microscope.[15] For example, myeloperoxidase (MPO) is a cytoplasmic protein highly expressed in the neutrophils of the myeloid lineage. Immunohistochemical staining for MPO can be used to detect leukemic blasts in AML patients[17]. Wright Giemsa Stains are also frequently used for

bone marrow smear.[3,6] The major components are methylene blue and its oxidative forms, azure A and azure B, and eosin.[18] They stain erythrocytes (RBCs) with pink, leukocyte nuclei with blue to purple and platelet cytoplasm with dark lilac.[18] The morphological images with Wright Giemsa staining of a normal and a B ALL bone marrow smear are shown in **Figure 2.1**.

However, having a sensitivity of 1-5% leukemic cells/WBCs [19,20] makes this test far from sensitive or discriminative to search for early stages of disease when the normal blood cell population is higher than 95% of the cell population. In addition, morphology possesses limited ability to discern key marker expression on leukemic cells to precisely distinguish leukemic cells from normal leukocytes or to discriminate leukemia-associated immunophenotypic abnormalities that can be used to guide therapeutic decisions.[21]

2.1.2.2 Flow Cytometry

Flow cytometry tests are often used to gain more insight into the disease state.[22] In flow cytometry, cells are incubated with fluorophore-conjugated antibodies that are specific to cell markers and are then hydrodynamically injected into a flow cell in a single line fashion and illuminated by a laser beam to secure information on protein markers, cell size and cell granularity.[23] The fluorescent dye emission are related to the cell markers and provide information on the absence or presence of certain markers and their expression level.[24] The most used markers for leukemic cells are cluster of differentiation (CD) markers. CD markers are a family of highly diverse membrane proteins mainly expressed on white blood cell surfaces or other immune system-related cells.[25]

Not only more descriptive than morphology, flow cytometry is also more sensitive with a sensitivity of 10^{-3} to 10^{-4} leukemic cells/WBCs.[20] However, this sensitivity is still insufficient for detecting low levels of leukemic cells in patients. Additionally, flow cytometric tests generally

require BM biopsies, which are highly invasive and painful. **Figure 2.2** displays the general workflow in conventional flow cytometry.

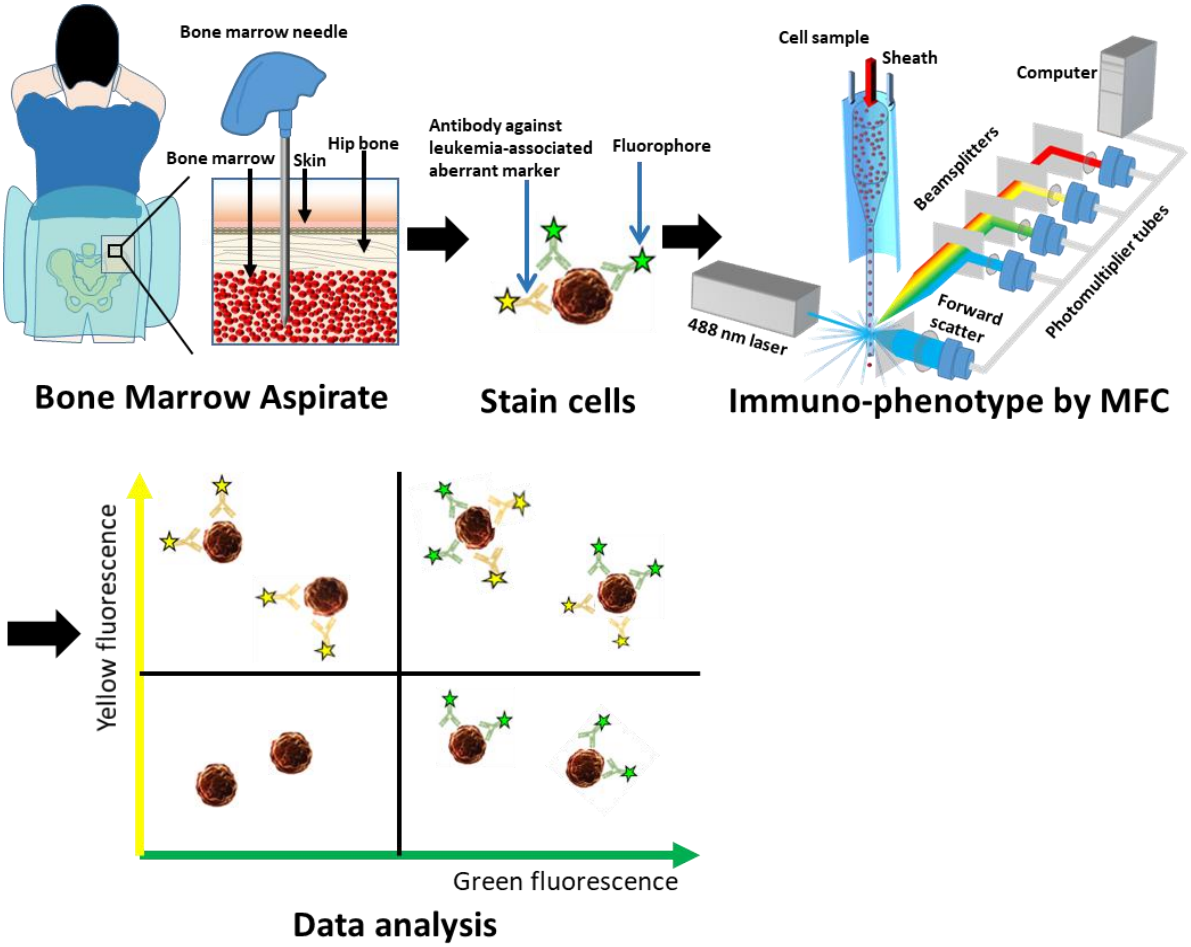


Figure 2.2 General workflow typically used in traditional flow cytometry, which includes taking bone marrow biopsy, staining the marrow cells, and introducing the cell sample into a flow cytometer for sample and data analysis.

2.1.2.3 Minimal Residual Disease

With current induction treatments, most leukemia patients have complete remission (CR), which is defined by the presence of less than 5% of leukemic blasts in the bone marrow through morphological assessment.[13,26,27] However, CR does not necessarily ensure positive patient outcome. The persistent leukemia cells after therapy, termed as minimal residual disease (MRD) [28], although low in numbers, may harbor drug resistance, reproduce to disease burden and cause

relapse[15,27]. Because the progression of leukemic MRD is associated with disease progression and serves as an important prognosticator to allow for tailoring treatments and predicting patient outcome, obtaining information on the numbers and phenotypic features of leukemic MRD is of paramount importance.[12,22,29] For example, when the MRD level cannot be properly monitored and similar maintenance therapies are given to all acute leukemia patients with CR, the non-targeted treatments may not be effective.[15] Specifically, the treatments may be overdose or unnecessary to patients with extremely low level MRD who are basically cured. On the other hand, the same treatments may be impotent or inappropriate to patients who have elevated level of MRD, making the treatment not efficacious preventing relapse. Only when the MRD level is precisely monitored in a timely manner can the appropriate patient-specific clinical actions be implemented.

However, the commonly used tests to monitor MRD lack the ability to pinpoint relapse from MRD due to their limited sensitivity. Hence, highly sensitive analytical methods are in high demand to meet this clinical need.

2.1.2.4 Detection of Leukemic Cells by Microfluidics

To address this challenge, our group previously developed a highly sensitive microfluidic assay to detect circulating leukemic cells (CLCs) from peripheral blood (PB) samples instead of BM for AML patients.[30] Briefly, 3 mL of PB from each AML patient was infused into three different sinusoidal microfluidic chips to affinity isolate three subgroups of leukemic cells. The enriched cells were then rinsed with buffer, stained with fluorochrome-linked antibodies specific to cell markers, released through cleavable linkers , and immunophenotyped by semi-automated fluorescence microscopy.[30] This assay greatly increased the sensitivity of leukemic cell detection in PB. Despite the heterogeneous leukemia-associated phenotypes in AML patients[31], the

microfluidic assay could detect 11-2684 leukemic cells/mL of blood for MRD patients. In addition, the assay obviated the requirement for an invasive BM biopsy, giving rise to higher sampling frequency. For instance, we could sample every week instead of every 2-3 months with bone marrow biopsy, allowing for closer monitoring of the disease. An illustration of this microfluidic assay is shown in **Figure 2.3**.

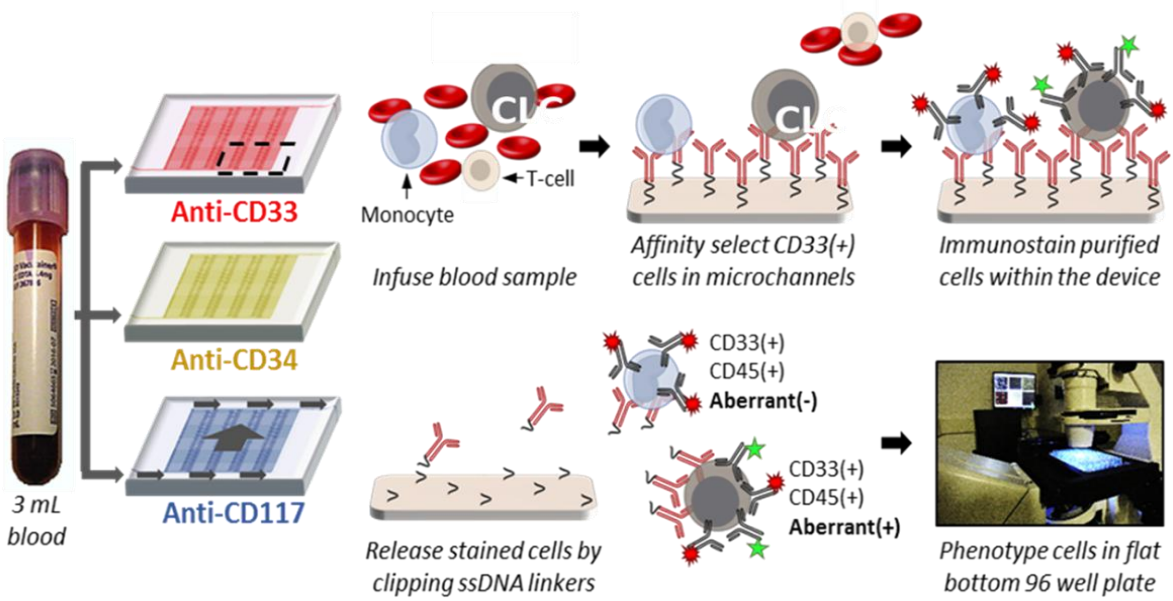


Figure 2.3 Detecting circulating leukemic cells (CLCs) from PB samples instead of BMs in AML patients. 3 mL of blood from each AML patient was perfused into three sinusoidal microfluidic chips to affinity isolate three subgroups of leukemic cells. The enriched cells were then rinsed with buffer, stained with fluorochrome-linked antibodies specific to cell markers, released through cleavable linkers, and immunophenotyped by semi-automated fluorescence microscopy. Reproduced from Ref. 30 with permission from The Royal Society of Chemistry

2.1.2.5 Immunophenotyping Microfluidic Chip-isolated Leukemic Cells by Fluorescence Microscopy

Fluorescence microscopy is a frequently used tool to immunophenotype cells to discern the presence of different markers.[30,32] The primary reason for the use of fluorescence microscopy is because it represents standard equipment in many labs[33], and possesses a relatively long history of use in clinical settings[34]. However, the data collection and processing normally require

multiple processing steps with various algorithms to generate accurate results, which eventually makes it challenging in terms of throughput.[33] This is particularly true when there are many cells in each sample or there are a lot of samples to be analyzed. For example, our AML microfluidic assay imaged the wells of microliter plates where the enriched leukemic cells were deposited following enzymatic release from the microfluidic chip. Due to the high number of cells and the semi-automated nature of the assay, the immunophenotyping of cells from a single chip required >8 h of processing time.

2.2 Flow Cytometry with Time Delayed Integration

2.2.1 Standard Flow Cytometry and Its Incompatibility with Microfluidic chip-isolated Cells

An alternative immunophenotyping method that possesses higher throughput compared to fluorescence microscopy is flow cytometry. Flow cytometry is a widely used tool for cell counting and immunophenotyping with higher automation capabilities compared to fluorescence microscopy.[35] Nevertheless, flow cytometry requires large numbers of cells to generate statistically reliable results. While it can process thousands of cells[36], in some cases up to 20,000 cells s^{-1} in modern flow cytometers[37], it does require at least thousands of cells to set up the proper gating levels and perform spectral compensation for channel cross-talk. However, in our microfluidic assay, the number of cells isolated from each microfluidic chip is normally in the range of several hundred to several thousand[30], far less than the amount of cells needed for standard flow cytometry to set up proper gates. Thus, the microfluidic chip-isolated cells can scarcely be analyzed by traditional flow cytometry. The spectral cross-talk in traditional flow cytometry is depicted in **Figure 2.4**.

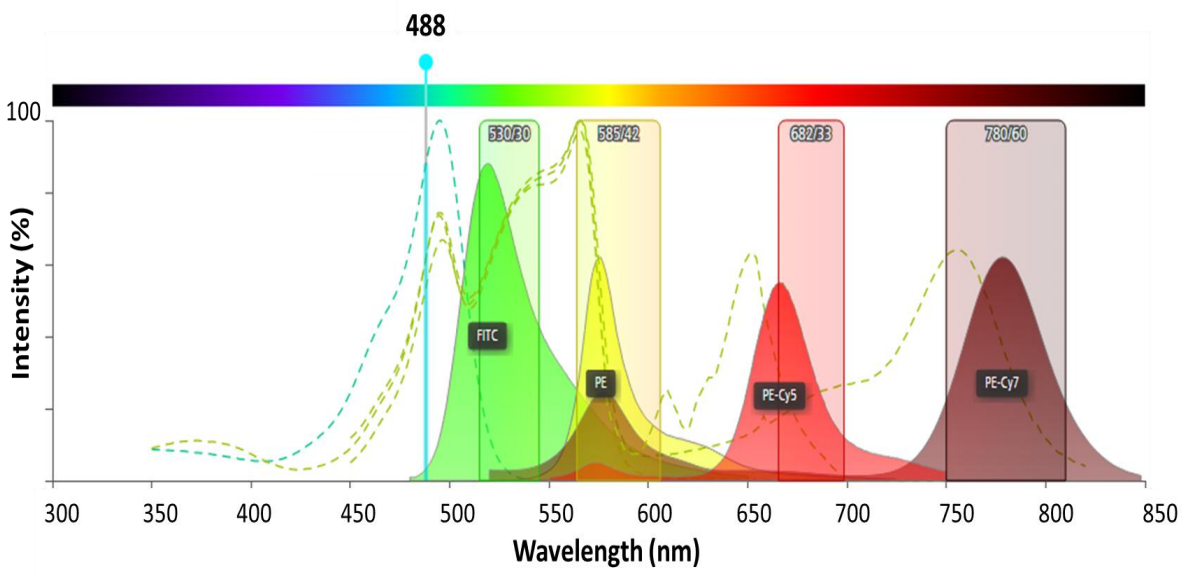


Figure 2.4 Spectral cross-talk in a standard flow cytometer with a 488 nm excitation light source. From left to right, the solid lines with color fillings are the fluorescence emission spectra of the four dyes FITC, PE, PE-Cy5 and PE-Cy7. The rectangles indicate the wavelength rages of the data collected by flow cytometry with bandpass filters. However, the data collected in each channel not only include the fluorescent photons from the dye of interest in that channel, but also the photons spilled over from other dyes due to spectral overlap. To tackle this issue, singly-stained control samples and experiments are necessary to perform spectral compensation. Figure simulated and reprinted from Fluorescence SpectraViewer on Thermo Fisher Scientific official website with permission[38].

2.2.2 Spectral Flow Cytometry with Time Delayed Integration

To address the gap between the time-consuming fluorescence microscopy and cell number-demanding flow cytometry, we present here a novel spectral flow cytometer with time-delayed integration (TDI). Instead of using filters for fluorescence color sorting, we employed a spectrograph so that the entire spectrum of the labeling dye could be interrogated so that spectral deconvolution could be employed to better identify the fluorescent labels attached to cell-resident molecules. This TDI spectral flow cytometer (TDI SFC) was designed to enumerate and immunophenotype chip-isolated leukemic cells from patients' PB samples. It should be noted that theoretically, not only leukemic cells but also any fluorescently stained cells can be analyzed using this TDI SFC.

2.2.2.1 Spectral Flow Cytometry vs Traditional Flow Cytometry

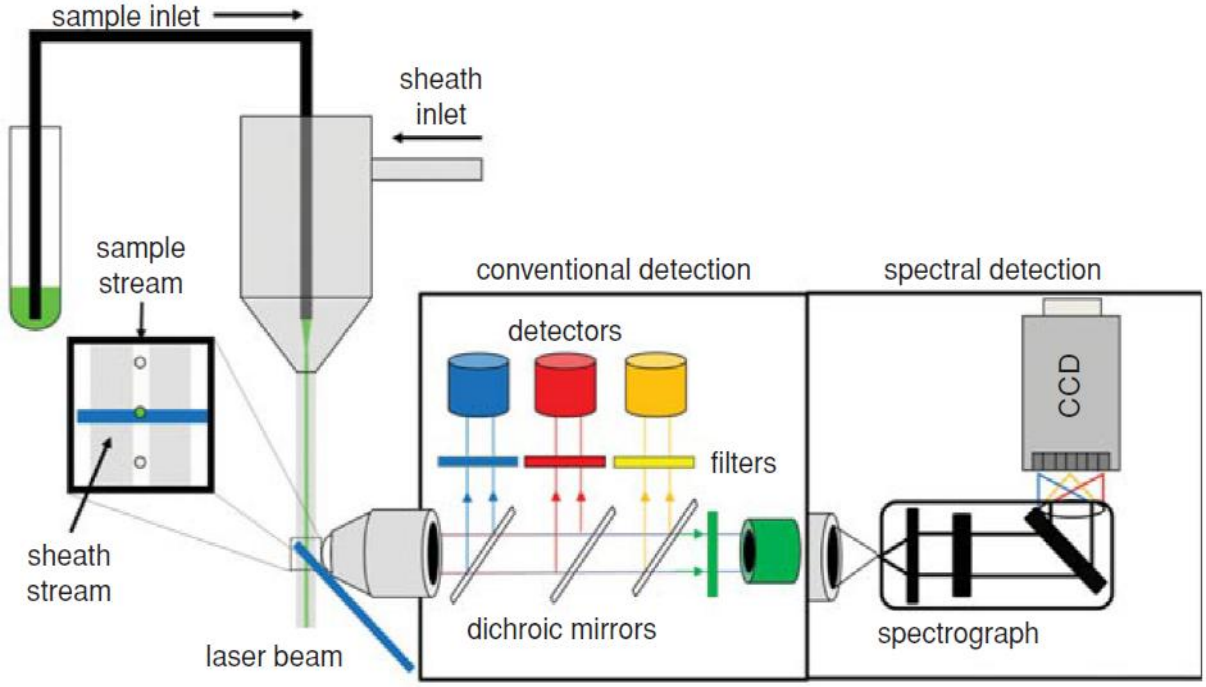


Figure 2.5 Fluorescence color-sorting differences between a traditional flow cytometer and a spectral flow cytometer. While the traditional flow cytometer uses dichroic mirrors and bandpass filters to collect light signals at certain wavelength ranges that are processed on separate photodetectors, the spectral flow cytometer utilizes dispersive optics, for example a spectrograph, to disperse light and collect the entire spectrum with the multi-color light processed on a CCD camera. Reproduced from Ref. 39 with permission. Copyright 2013 John Wiley and Sons.

The primary difference between a traditional flow cytometer and a spectral flow cytometer resides within the optics and the photon signal they collect. While the traditional flow cytometer uses dichroic mirrors and bandpass filters to collect light signals at certain wavelength ranges, the spectral flow cytometer utilizes dispersive optics, for example, a spectrograph, to disperse light and collect the entire fluorescence emission spectrum.[39] Subsequently, spectral deconvolution is usually implemented to resolve each individual spectrum from the total spectrum. Although the two cytometers may share the same fluidic system, spectral flow cytometry allows for higher usage of the fluorescence emission because the entire emission spectrum is recorded. This results in advantages over traditional flow cytometry. For instance, in

traditional flow cytometry, care must be taken when using dyes that exhibit significant spectral overlap. But, using these dyes is not problematic in spectral flow cytometry because spectral deconvolution is performed to resolve the individual spectrum from each dye.[40] The differences in the color sorting optical trains between a traditional flow cytometer and a spectral flow cytometer is shown in **Figure 2.5**.

2.2.2.2 TDI mode: How It Works and What It Can Do

Our TDI SFC utilizes a Charge-Coupled-Device (CCD) photon detector, distinguishing it from a traditional flow cytometer, which uses photomultiplier tubes (PMTs) or/and photodiodes (PDs). In our case, the CCD is programmed to operate in a time-delayed integration (TDI) mode. TDI has been widely used in remote imaging systems to acquire high resolution images of moving objects with weak light.[41] However, the TDI mode of CCD operation has only rarely been applied in bioanalytical chemistry.[42,43] Briefly, the advantage of using the TDI mode as compared to a snapshot CCD operational mode is: (1) The duty cycle of TDI is much better than the snapshot mode because in TDI, the shutter remains open during the entire experimental run; and (2) the signal-to-noise ratio (SNR) for TDI is higher than the snapshot mode with the gain in SNR approximately equal to $n^{1/2}$, where n is the number of CCD pixels in a single column.

On a CCD image sensor, multiple pixels collect light signals from an object and transduce them into charge signals. By applying proper electrical conditions, the charge signals are simultaneously shifted across the pixel rows (also known as parallel registers) and go to the serial register to be read out.[44] TDI mode eliminates the need for a shutter and takes advantage of the signal shifting process.[45] Namely, the object's linear velocity and the CCD's parallel shift rate or readout rate are synchronized so that the photoelectrical signal will accumulate in only a few pixel rows at readout, increasing the SNR.[41] With TDI SFC, each pixel row records a

spectrum, while the pixel column contains the single cell signal. A schematic of a fluorescently labeled cell being processed using the TDI SFC is shown in **Figure 2.6**.

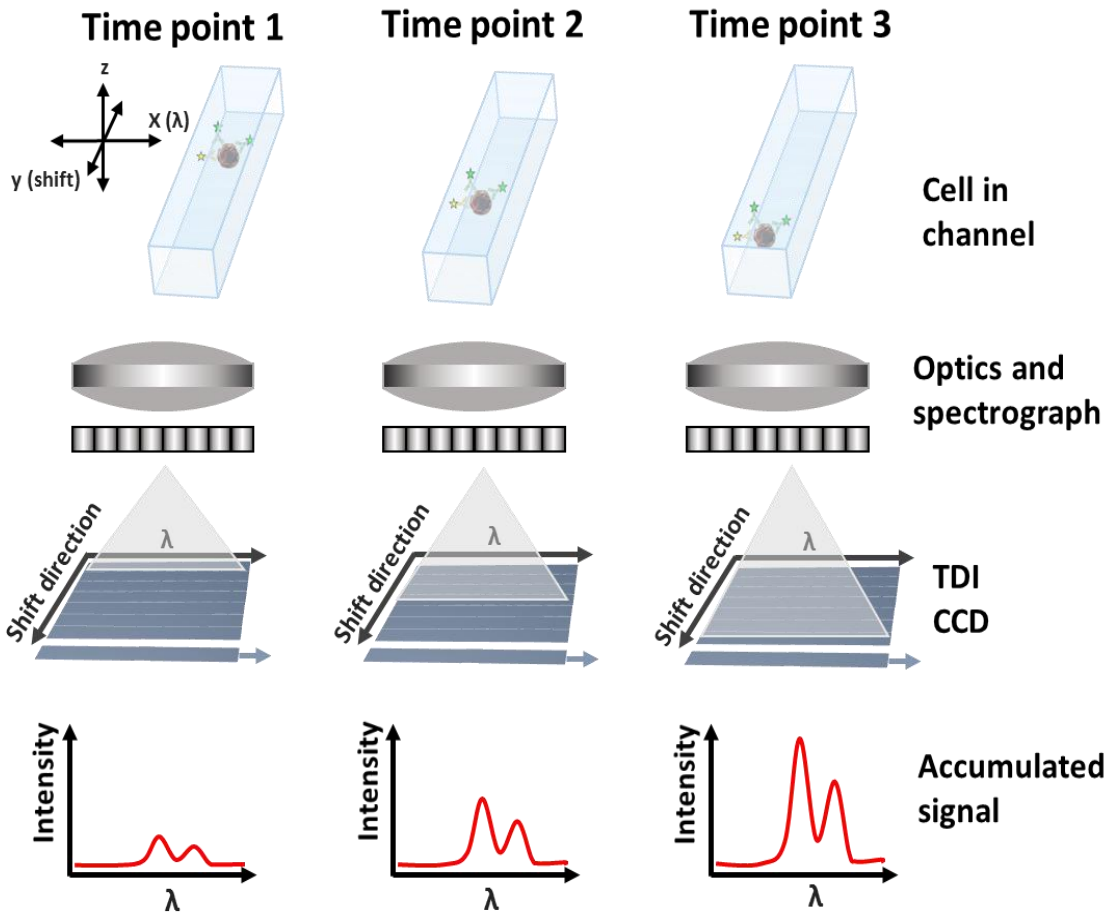


Figure 2.6 Schematic detecting a fluorophore-labeled cell in our TDI SFC. Because the cell's linear velocity and the CCD's readout rate are synchronized, as the cell moving forward on the shift axis, the fluorescent signal is transported and integrated within only a few pixel rows at readout. Each pixel row on spectral direction collects a full spectrum due to the spectrum dispersion by the spectrograph.

2.2.2.3 Critical Factors when Operating TDI SFC

2.2.2.3.1 Synchronization of Particle Moving with CCD Readout

Due to the unique features of the TDI SFC, there are several factors where attention should be paid to facilitate successful experiments. First and foremost, it is critical to synchronize the linear velocity of the fluorescent particle with the pixel row shift rate of the CCD.

Only in this case can the signal be accumulated within a few pixel rows and optimal SNR be obtained.[42] If the particle's linear velocity is not synchronized well with the CCD's pixel row shift rate, there will be image blurring and the photoelectrical signal will be distributed over multiple pixel rows, reducing the SNR.

2.2.2.3.2 Focusing of Particles

It is widely accepted that in microfluidic channels with pressure-driven flow, due to the flow's parabolic nature dictated by Reynold numbers (Re) <2000, the linear velocities of the fluorescent particles near to the channel wall are lower than that at the centerline.[46,47] The parabolic profile of the flow in microfluidic channels is illustrated in **Figure 2.7**.

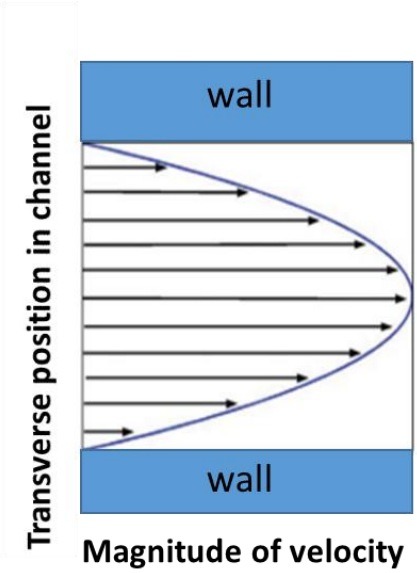


Figure 2.7 The parabolic flow profile in microfluidic channels. In typical microfluidic channels where Re is far below 2000, the fluid's linear velocity at the wall is lower than that at the channel centerline. It results in position-dependent non-uniformity of linear velocities in the channel. Adapted from Ref. 47 with permission from The Royal Society of Chemistry.

The equation to calculate Re can be expressed by equation 2.1:

$$Re = \frac{\rho u L}{\mu} \qquad \text{(Eq 2.1)}$$

where ρ is the density of the fluid (kg/m^3), u is the velocity of the fluid with respect to the object (m/s), L is a characteristic linear dimension (m) of the channel and μ is the dynamic viscosity of

the fluid (Pa·s or N·s/m^2 or kg/m·s). This is exacerbated by the no-slip condition, where the velocity at the wall is zero and thus, giving rise to Taylor dispersion.

Usually, the CCD's parallel shift rate is constant during an experiment. However, due to the flow's parabolic nature in the microfluidic channel, the fluorescently labeled analytes in the flow possess various linear velocities, making it difficult to synchronize the analytes' velocities with the CCD's parallel shift rate. Hence, this nonuniformity of the particles' linear velocities is unwanted if optimal synchronization is desired.

A strategy to obviate this issue is to use hydrodynamic flow focusing, in which a sheath flow is used to surround the sample flow. Because the sheath is operated at a higher velocity than the sample stream velocity, the sample velocity is accelerated and thus, focuses to the centerline of the flow where the velocity becomes more constant, and thus can match the parallel shift rate of the CCD with the linear velocity of the biological cells to be interrogated. The added benefit is that the cells also align themselves into a single file so that each cell can be processed individually; the sampling efficiency for single cells increases dramatically compared to non-focused flow.

2.3 Flow Device for TDI SFC

2.3.1 T Shaped Microfluidic Chip

To facilitate the synchronization of the analytes with the CCD's parallel shift rate in TDI, a microfluidic flow cell adopting a simple T shape has been fabricated to hydrodynamically focus the analytes within the microfluidic channel. This flow cell is capable of confining the fluorescent particles into a relatively narrow cross-area at the channel center and enhancing the uniformity of their linear velocities. In terms of fabricating the microfluidic flow cell, poly (methylmethacrylate), PMMA, a cheap and easily accessible plastic, was used as the substrate.

The fact that PMMA possess excellent optical transparency[48,49], relatively low autofluorescence[48] and its refractive index (RI) is very close to glass (PMMA RI=1.50[49,50] and glass RI=1.51[51] at 488 nm, respectively) made it an ideal substrate for the TDI SFC application. Additionally, PMMA has minimal water absorption, greatly reducing the possibility of channel deformation after multiple times of usage. Channel deformation is detrimental to TDI because it can cause changes of the analytes' linear velocities, and thus should be minimized.

2.3.2 Square Capillary

To confirm the benefit of hydrodynamic focusing in the microfluidic flow cell, a square capillary with no focusing was used as a flow device to test the operation of the TDI SFC using fluorescent beads as biological cell models. The data obtained in each flow device are compared. The differences of no focusing in a square capillary versus 1D focusing in a T-shaped microchip flow cell are shown in **Figure 2.8**. More detailed data and COMSOL simulations are in Chapter 4.

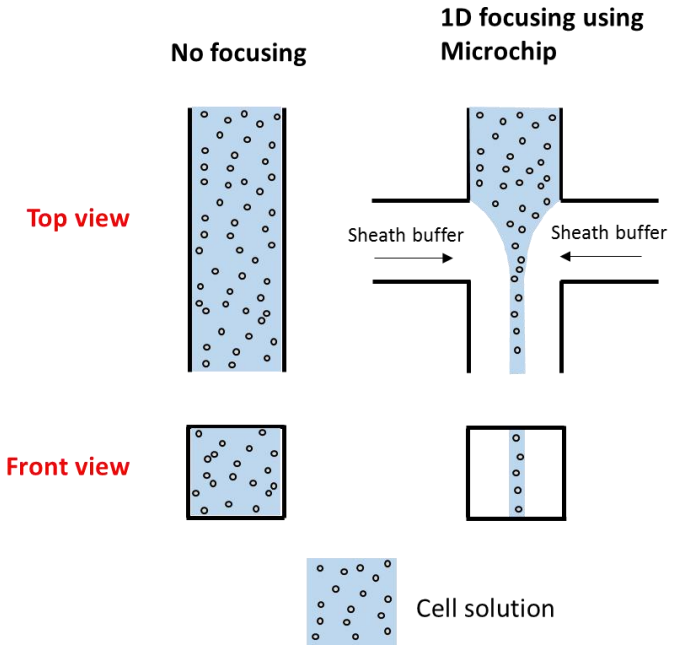


Figure 2.8 Flow devices used in this thesis. The T-shaped microfluidic flow device is capable of providing 1D flow focusing of the analytes, improving the uniformity of the analytes' linear velocities for optimal synchronization with the TDI CCD's parallel shift rate.

2.4 Conclusion

Leukemia is a serious malignancy that necessitates more sensitive detection methods that are able to use PB samples instead of BM samples to facilitate more frequent monitoring of disease recurrence, which can result in a better outcome for patients, especially those with acute type leukemias.[2] Our group has previously reported the detection of rare leukemic cells in acute myeloid leukemia (AML) patients' blood samples using affinity isolation through curvilinear microfluidic chips and cell immunophenotyping via fluorescence microscopy.[30] Despite the heterogeneity of the leukemia-associated cell immunophenotypes in AML patients, our microfluidic assay improved the sensitivity of leukemic blast detection from 10^{-3} – 10^{-4} using conventional flow cytometry and 1-5% using morphology to 10^{-6} (leukemic cells/WBCs).[13,20,30]

However, immunophenotyping the chip-isolated leukemic cells by fluorescence microscopy often required long and laborious processing with various data analysis algorithms to generate reliable results and therefore can be throughput-inefficient. Flow cytometry is another immunophenotyping method with higher throughput than fluorescence microscopy and is often used to analyze leukemic cells from leukemia patients [35]; it can process cells rapidly (10^4 cells/s), but requires a large number of cells to generate statistically reliable results; and due to the transient nature of the cells within the sampling zone, it becomes difficult to detect cells with low antigen expression levels. Additionally, because traditional flow cytometers are bandpass filter-based, they only acquire data at certain wavelengths instead of acquiring the entire emission spectra, possibly resulting in bias and/or misclassification, especially when there are only limited numbers of cells or the cells are generating considerable autofluorescence due to disease progression or treatment. Therefore, for chip-isolated cells that are usually in the range of 10,000 or below, traditional flow cytometry is not compatible for accurate cell immunophenotyping. To

address the issues between the time-consuming fluorescence microscopy and the cell number-demanding flow cytometry, we will report on a spectral flow cytometer with time delayed integration (TDI SFC).

2.5 Summary of Subsequent Chapters

Chapter 3 describes the detection and analysis of glycopeptides by a semi-targeted data acquisition strategy, the inclusion approach, as opposed to the top 10 approach in the commonly employed DDA mode during MS/MS experiments. In this study, to facilitate the inclusion approach, a publicly accessible software tool GlycoPep MassList is developed to aid in generating large numbers of glycopeptide masses. To validate the advantages of the inclusion approach, human IgG, an important endogenous glycoprotein of the immune system, is digested by trypsin and the resulting peptides and glycopeptides are analyzed by mass spectrometry using inclusion and top 10. A more complex sample, a mixture of bovine fetuin and chicken avidin is also tested with inclusion and top 10. The data are manually analyzed and the glycopeptide coverages are summarized to demonstrate the value of the software tool GlycoPep MassList and the merits of the inclusion approach as a MS/MS data acquisition strategy.

Chapter 4 introduces the characterization of an in-house developed time delayed integration spectral flow cytometer (TDI SFC) designed for enumerating and immunophenotyping microfluidic chip-isolated cells, in particular chip-affinity selected leukemic cells. In this study, the TDI SFC is first characterized using fluorescent calibration beads. The field of view and readout time of the TDI SFC are measured to find out the theoretical ideal linear velocity of the analytes for optimal clocking with respect to the TDI CCD. Because TDI depends on the synchronization of the linear velocity of the fluorescent particles and the readout rate of the TDI CCD, fluorescent calibration beads are used as cell models to

discover the optimal flow rates for ideal synchronization. COMSOL simulations are carried out to confirm the agreement of the theoretical results and experimental results. Subsequently, optimizations in various aspects, such as using higher power light source, applying a higher magnification objective and increasing the total full frame integration time, are performed to improve the fluorescence intensity and SNR. The detection efficiency and throughput were also evaluated by comparing the data from our TDI SFC with the data from a commercial flow cytometer.

Chapter 5 details the validation of the TDI SFC by successfully detecting stained leukemic cell lines. The stained B cell acute lymphoblastic leukemia (B-ALL) cells, SUP-B15 cell lines, were used as CLC models to validate the application of the TDI SFC. The detection markers consisted of a nuclear stain 7-aminoactinomycine D (7-AAD) and anti-Terminal-deoxynucleotidyl Transferase (TdT) FITC. 7-AAD as a nuclear stain is used to make sure that the events detected are truly from a cell and not from an interferent. Terminal-deoxynucleotidyl transferase (TdT) is a B-ALL marker that is positively expressed in ~90% of ALL patients and a commonly included marker for B-ALL leukemic cells detection by flow cytometry. Thus, it is used as another detection marker. To perform spectral deconvolution, unmixing algorithms were developed and optimized to resolve each marker's fluorescence emission from to total emission spectra acquired by the TDI SFC.

Chapter 6 outlines the future work that can potentially push the TDI SFC into new areas of application with better performance compared to traditional flow cytometry and also improve the performance of the reported TDI SFC. A critical aspect will be replacing the one dimensional (1D) focusing microchip reported in Chapter 4 with a two dimensional (2D) focusing chip. The 2D focusing is capable of confining the cells into a very small cross section at the center of the

channel. Therefore, it results in two main advantages. First, with 2D focusing the cells' linear velocities will be much more uniform than that in the T-shaped microfluidic flow cell with 1D focusing, leading to much higher percentage of the cells with velocities that more closely match the parallel shift rate of the TDI CCD. Second, the variance of the fluorescence signal from the cells is primarily due to the marker expression level differences on the cells instead of the positional difference of the cells at the microfluidic channel. Another modification that can improve the performance of the system is to improve the CCD camera's frame rate. If the frame rate is faster, the cells can move faster in the microfluidic channel. This can not only improve the throughput and make data collection more time-efficient, but also can reduce the probability of the cells' sedimentation in the fluidic system and obtain higher detection efficiency.

2.6 References

(1) Orkin, S. H.; Zon, L. I. *Cell* **2008**, *132*, 631-644.

(2) Jordan, C. T. *Current Opinion in Cell Biology* **2004**, *16*, 708-712.

(3) Sharma, R.; Nalepa, G. *Pediatrics in review* **2016**, *37*, 101-111.

(4) Udensi, U. K.; Tchounwou, P. B. *J Exp Clin Cancer Res* **2014**, *33*, 106.

(5) Uchiyama, T.; Yodoi, J.; Sagawa, K.; Takatsuki, K.; Uchino, H. *Blood* **1977**, *50*, 481-492.

(6) Hayne, C. C.; Winer, E.; Williams, T.; Chaves, F.; Khorsand, J.; Mark, H. F. *Exp Mol Pathol* **2006**, *81*, 62-71.

(7) Melo, J. V.; Robinson, D. S. F.; Deoliveira, M. P.; Thompson, I. W.; Lampert, I. A.; Ng, J. P.; Galton, D. A. G.; Catovsky, D. *Journal of Clinical Pathology* **1988**, *41*, 951-959.

(8) Borowitz, M. J.; Devidas, M.; Hunger, S. P.; Bowman, W. P.; Carroll, A. J.; Carroll, W. L.; Linda, S.; Martin, P. L.; Pullen, D. J.; Viswanatha, D.; Willman, C. L.; Winick, N.; Camitta, B. M.; Childrens Oncology, G. *Blood* **2008**, *111*, 5477-5485.

(9) Buccisano, F.; Maurillo, L.; Del Principe, M. I.; Del Poeta, G.; Sconocchia, G.; Lo-Coco, F.; Arcese, W.; Amadori, S.; Venditti, A. *Blood* **2012**, *119*, 332-341.

(10) Kato, M.; Manabe, A. *Pediatrics International* **2018**, *60*, 4-12.

(11) Rowe, J. M.; Buck, G.; Burnett, A. K.; Chopra, R.; Wiernik, P. H.; Richards, S. M.; Lazarus, H. M.; Franklin, I. M.; Litzow, M. R.; Ciobanu, N. *Blood* **2005**, *106*, 3760-3767.

(12) Weng, X. Q.; Shen, Y.; Sheng, Y.; Chen, B.; Wang, J. H.; Li, J. M.; Mi, J. Q.; Chen, Q. S.; Zhu, Y. M.; Jiang, C. L.; Yan, H.; Zhang, X. X.; Huang, T.; Zhu, Z.; Chen, Z.; Chen, S. J. *Blood Cancer J* **2013**, *3*, 1-9.

(13) Buckley, S. A.; Appelbaum, F. R.; Walter, R. B. *Bone Marrow Transplant* **2013**, *48*, 630-641.

(14) Meshinchi, S.; Thomson, B.; Finn, L. S.; Leisenring, W.; Green, C.; Radich, J. P.; Loken, M.; Hawkins, D. *Journal of Pediatric Hematology/Oncology* **2001**, *23*, 585-590.

(15) Ryan, D. H.; van Dongen, J. J. In *Immunologic approaches to the classification and management of lymphomas and leukemias*; Springer, 1988, pp 173-207.

(16) Campana, D. *Hematology-Oncology Clinics of North America* **2009**, *23*, 1083-1098.

(17) Pinkus, G. S.; Pinkus, J. L. *Modern Pathology* **1991**, *4*, 733-741.

(18) Dunning, K.; Safo, A. O. *Biotech Histochem* **2011**, *86*, 69-75.

(19) Coustan-Smith, E.; Behm, F. G.; Sanchez, J.; Boyett, J. M.; Hancock, M. L.; Raimondi, S. C.; Rubnitz, J. E.; Rivera, G. K.; Sandlund, J. T.; Pui, C. H.; Campana, D. *Lancet* **1998**, *351*, 550-554.

(20) Al-Mawali, A.; Gillis, D.; Lewis, I. *American Journal of Clinical Pathology* **2009**, *131*, 16-26.

(21) Wood, B. L. *Journal of Hematopathology* **2015**, *8*, 191-199.

(22) Ossenkoppele, G. J.; van de Loosdrecht, A. A.; Schuurhuis, G. J. *British Journal of Haematology* **2011**, *153*, 421-436.

(23) Picot, J.; Guerin, C. L.; Le Van Kim, C.; Boulanger, C. M. *Cytotechnology* **2012**, *64*, 109-130.

(24) Cossarizza, A.; Chang, H. D.; Radbruch, A.; Akdis, M.; Andra, I.; Annunziato, F.; Bacher, P.; Barnaba, V.; Battistini, L.; Bauer, W. M.; Baumgart, S.; Becher, B.; Beisker, W.; Berek, C.; Blanco, A.; Borsellino, G.; Boulais, P. E.; Brinkman, R. R.; Buscher, M.; Busch, D. H., et al. *European Journal of Immunology* **2017**, *47*, 1584-1797.

(25) Engel, P.; Boumsell, L.; Balderas, R.; Bensussan, A.; Gattei, V.; Horejsi, V.; Jin, B. Q.; Malavasi, F.; Mortari, F.; Schwartz-Albiez, R.; Stockinger, H.; van Zelm, M. C.; Zola, H.; Clark, G. *J Immunol* **2015**, *195*, 4555-4563.

(26) Bielorai, B.; Golan, H.; Trakhtenbrot, L.; Reichart, M.; Toren, A.; Daniely, M.; Zilberstein, Y.; Amariglio, N.; Rechavi, G.; Kaplinsky, C. *Cancer Genetics and Cytogenetics* **2002**, *138*, 64-68.

(27) Bjorklund, E.; Mazur, J.; Soderhall, S.; Porwit-MacDonald, A. *Leukemia* **2003**, *17*, 138-148.

(28) Cruz, N. M.; Mencia-Trinchant, N.; Hassane, D. C.; Guzman, M. L. *International Journal of Laboratory Hematology* **2017**, *39*, 53-60.

(29) Patel, B.; Foroni, L. *Clinical Leukemia* **2006**, *1*, 108-117.

(30) Jackson, J. M.; Taylor, J. B.; Witek, M. A.; Hunsucker, S. A.; Waugh, J. P.; Fedoriw, Y.; Shea, T. C.; Soper, S. A.; Armistead, P. M. *Analyst* **2016**, *141*, 640-651.

(31) Kern, W.; Bacher, U.; Haferlach, C.; Schnittger, S.; Haferlach, T. *Best Practice & Research Clinical Haematology* **2010**, *23*, 379-390.

(32) Shi, W. T.; Wang, S. Q.; Maarouf, A.; Uhl, C. G.; He, R.; Yunus, D.; Liu, Y. L. *Lab on a Chip* **2017**, *17*, 3291-3299.

(33) Marjanovic, I.; Kanduser, M.; Miklavcic, D.; Keber, M. M.; Pavlin, M. *Journal of Membrane Biology* **2014**, *247*, 1259-1267.

(34) Takehara, H.; Kazutaka, O.; Haruta, M.; Noda, T.; Sasagawa, K.; Tokuda, T.; Ohta, J. *Aip Advances* **2017**, *7*, 1-8.

(35) Muratori, M.; Forti, G.; Baldi, E. *Cytometry A* **2008**, *73*, 785-787.

(36) Shapiro, H. M.; Perlmutter, N. G. *Cytometry Part A* **2006**, *69A*, 620-630.

(37) Myung, J. H.; Hong, S. *Lab on a Chip* **2015**, *15*, 4500-4511.

(38) Thermo Fisher Scientific Fluorescence SpectraViewer, 2018.

(39) Nolan, J. P.; Condello, D. *Curr Protoc Cytom* **2013**, *Chapter 1*, 1-13.

(40) Nolan, J. P. *Cytometry Part A* **2014**, *85A*, 10-11.

(41) Han, L. Q.; Yao, S. Y.; Xu, J. T.; Xu, C. *Microelectronics Reliability* **2013**, *53*, 400-404.

(42) Emory, J. M.; Soper, S. A. *Analytical Chemistry* **2008**, *80*, 3897-3903.

(43) Sweedler, J. V.; Shear, J. B.; Fishman, H. A.; Zare, R. N.; Scheller, R. H. *Analytical Chemistry* **1991**, *63*, 496-502.

(44) Aikens, R. S.; Agard, D. A.; Sedat, J. W. *Methods in Cell Biology* **1989**, *29*, 291-313.

(45) Bastian, U.; Biermann, M. *Astronomy & Astrophysics* **2005**, *438*, 745-755.

(46) Stone, H. A.; Kim, S. *Aiche Journal* **2001**, *47*, 1250-1254.

(47) Pennathur, S. *Lab on a Chip* **2008**, *8*, 383-387.

(48) Hayashi, M.; Hattori, A.; Kim, H.; Terazono, H.; Kaneko, T.; Yasuda, K. *International Journal of Molecular Sciences* **2011**, *12*, 3618-3634.

(49) Thakkar, I. G.; Lear, K. L.; Vickers, J.; Heinze, B. C.; Reardon, K. F. *Lab on a Chip* **2013**, *13*, 4775-4783.

(50) Piruska, A.; Nikcevic, I.; Lee, S. H.; Ahn, C.; Heineman, W. R.; Limbach, P. A.; Seliskar, C. J. *Lab on a Chip* **2005**, *5*, 1348-1354.

(51) Klar, T. A.; Jakobs, S.; Dyba, M.; Egner, A.; Hell, S. W. *Proceedings of the National Academy of Sciences of the United States of America* **2000**, *97*, 8206-8210.

CHAPTER III

GlycoPep MassList: Software to Generate Massive Inclusion Lists for Glycopeptide Analyses

This work has been published by the journal Analytical and Bioanalytical Chemistry, with reprint permission from the journal.

3.1 Introduction

Protein glycosylation is one of the most significant and fundamental post translational modifications (PTMs) in nature.[1] Glycosylation, like other PTMs, leads to the increased diversification of protein structures and functions.[2,3] The most common type of glycosylation, N-linked glycosylation, occurs when the glycan is appended to the side chain of a asparagine following the consensus sequence N-X-T/S, where X is any amino acid except proline.[4] The glycans appended at N-linked glycosylation sites mainly depend on the glycosyltransferase enzyme availability and the microenvironment of the protein.[5,6] The glycans play crucial roles in a variety of biological processes including protein folding[7], protein stabilization[8], immune response[9], cell-environment communication[10], and fertilization[11]. Many times, a particular glycan profile is essential for the glycoprotein to function optimally.[9] Alterations in the glycosylation site, the extent of glycosylation, or the glycosylation profile are associated with a broad spectrum of diseases[12,13], ranging from rheumatoid arthritis[14], Alzheimer's disease[15] to prostate[16], colorectal[17] and breast cancer[18]. The identification of abnormally glycosylated proteins has biomedical value because these features may serve as disease biomarkers.[19-21] Therefore, in order to gain more understanding in protein structure-function correlations and/or to exploit

glycoproteins in biomarker discovery and disease diagnosis, one must be able to characterize the glycosylation on N-linked proteins in an efficient manner.

Analysis of N-linked glycosylation can be done in a variety of ways, but often researchers prefer to obtain the relevant information by analyses of proteolyzed glycopeptides.[22,23] In these experiments, the glycans remain linked to the site on the protein at which they reside. In this case, the most common workflow involves tryptic digestion of the protein, followed by LC-MS/MS analysis.[24] Past research has demonstrated that in all but the simplest cases, high resolution MS data are not sufficient to accurately identify glycopeptides.[25] MS/MS is necessary to distinguish among various possible glycopeptide assignments for any given MS peak.[22,25] The MS/MS data can be used to determine the glycosylated protein sequence, the glycosylation site, and the glycan composition.[26,27]

Collision induced dissociation (CID) is the most commonly used dissociation method for glycopeptide analysis. One of its key advantages is its rapid duty cycle, allowing for data acquisition on multiple co-eluting glycopeptides. However, glycopeptide analysis by CID is still challenging, due to glycopeptides' intrinsic low abundance. The macroheterogeneity which comes from the difference of glycosylation site occupancy, and the microheterogeneity, which results from the attachment of different glycans to one specific glycosylation site, render each glycoform in a low copy number after proteolysis.[28-30] Consequently, the most commonly used MS/MS data-dependent acquisition (DDA) mode, where the most intense ions in the full MS scan are selected for MS/MS, is not optimal for glycopeptide analyses because the high-abundant precursor ions, which are often non-glycosylated peptides, are redundantly selected for MS/MS, while the relatively low abundant glycopeptide ions may not trigger MS/MS, even if dynamic exclusion is enabled.[31] This experiment results in missed detection of glycopeptides and limited

glycopeptide coverage, particularly when the sample has a complex matrix that brings high background corresponding to non-glycosylated peptides.

A significant thrust of research in the area of glycopeptide analysis is, therefore, focusing on the problem of enhancing the number of glycopeptides selected for MS/MS analysis in a given sample. Sample preparation strategies, particularly glycopeptide enrichment, can contribute to this solution by reducing the number of non-glycosylated peptides present in the sample.[32,33] New MS methods are also needed. One such strategy that has not yet been readily adopted, but which theoretically could benefit the field, is a targeted analysis approach, taking advantage of instruments' ability to selectively conduct MS/MS on ions preloaded onto an inclusion list.

Targeted data acquisition is a well-known strategy that has been used to alleviate the biased ion selection inherent in DDA strategies in other fields, but it has not yet been commonly applied to the field of glycoproteomics. The proteomics field has already demonstrated that targeted data acquisition consumes fewer MS/MS scans on peptides that have high abundance but are not of interest to the investigators.[31] Targeted data acquisition strategies have been successfully employed to study arginine methylation[34], for example. The method, however, does not readily transfer to the field of glycopeptide analysis, primarily because the glycan component on the glycopeptide may be any one of hundreds of different masses. With such a tremendous variety of glycans, it would be a time-consuming process to generate an inclusion glycopeptide mass list every time a new glycoprotein is to be studied.

At least two groups have shown that using inclusions lists for glycopeptide analysis is a promising approach. Yin Wu and colleagues developed a software-based strategy that adds putative glycopeptide ions to an inclusion list for targeted MS/MS experiments.[35] Their strategy

relies on using the software GlycoPID to identify glycopeptides in untargeted manner first, then additional ions in the high resolution spectra that may also be glycopeptides are added to an inclusion list on a second, or third, or fourth round of experiments. This iterative fashion of targeted MS/MS experiment clearly achieves enhanced coverage, compared to a single round of untargeted experiments. The disadvantage of this approach is that multiple LC-MS/MS analyses are required for each sample, and the GlycoPID tool must be used for the glycopeptide assignments. GlycoPID is just one of many emerging bioinformatics platforms used for the glycopeptide analysis, and often users may want to use different software (or even manual analysis) to analyze their data. More recently, Froehlich and coworkers developed a mass defect classifier that tentatively identifies potential glycopeptide ions in a first-pass LC-MS analysis, and those ions could then be loaded onto an inclusion list during a re-analysis experiment.[36] This approach may be advantageous to the one built into GlycoPID, because no glycopeptide compositions need to be assigned initially, and users can select any glycopeptide analysis software they choose. This approach still requires an initial data acquisition phase, followed by data analysis, followed by at least one more round of data acquisition.

In the work described herein, we provide a software tool designed to generate inclusion lists for glycopeptides prior to any data acquisition steps. This software allows users to target glycopeptides for MS/MS in a single round of experiments, reducing the sample requirements and analysis time that results from doing multiple LC-MS injections. Any glycopeptide analysis software can be used in conjunction with this tool to interpret the data. The software was designed such that users could input any protein sequence of interest, and the peptides containing N-linked glycosylation sites are coupled to an on-board glycan library, which can be customized by the user, producing appropriate glycopeptide masses that can easily be uploaded into an

inclusion list for MS/MS analyses. A preliminary demonstration of the potential application of the software is described, along with some strategies that can potentially maximize the benefit of identifying glycopeptides using this approach.

3.2 Materials and Methods

3.2.1 Materials and Reagents

Fetuin from fetal bovine serum, avidin from chicken egg white, and human serum IgG were purchased from Sigma Aldrich (St. Louis, MO). Sequencing grade trypsin was obtained from Promega (Madison, WI). HPLC grade acetonitrile and methanol, ammonium bicarbonate, urea, guanidine hydrochloride (GdnHCl), dithiothreitol (DTT), iodoacetamide (IAM), and formic acid were purchased from Sigma Aldrich (St. Louis, MO). All the reagents were of analytical grade or better and were used without further purification.

3.2.2 Protein Digestion

Between 100 and 400 μg bovine fetuin or chicken avidin was dissolved in 50 mM ammonium bicarbonate buffer (pH 8.0) and denatured by adding urea until the concentration reached 6 M. The disulfides of the denatured proteins were reduced by DTT, which was added to reach a final concentration of 10 mM. The DTT reacted for one hour at room temperature. IAM was added to alkylate the disulfides; its final concentration was 25 mM. It reacted for one hour in the dark at room temperature. The excess IAM was quenched by DTT (final concentration 30 mM) for half an hour at room temperature. Subsequently, the protein solutions were diluted with 50 mM ammonium bicarbonate buffer to reach a urea concentration of 1 M, prior to incubation with trypsin (trypsin/protein, 1/30) at 37°C for 20 hours.

Human IgG, 100 μg, was dissolved in 50 mM ammonium bicarbonate buffer (pH 8.0) and was denatured by the addition of GdnHCl until the solution concentration reached 6 M. The

reduction, alkylation and quenching of excess IAM were the same as described above. The resulting protein solution was buffer exchanged with 50 mM ammonium bicarbonate buffer for two times to remove most of the GdnHCl. The buffer exchanged solution was made to a final volume of 100 µL before incubation with trypsin at a trypsin/protein ratio of 1/30 at 37°C for 20 hours. Finally, the digestion was stopped by adding formic acid with a formic acid/digestion solution ratio of 1/100. Each digested protein sample was aliquoted and stored at -20°C until it was analyzed.

3.2.3 LCMS

Glycoprotein samples were separated on a reverse phase C18 capillary column (300 µm i.d. × 5 cm, 100 Å pore size, Micro-Tech, Vista, CA) online using a Waters Acquity high performance liquid chromatography (Milford, MA) prior to mass spectrometric analysis in an LTQ Orbitrap Velos Pro hybrid mass spectrometer (Thermo Scientific, San Jose, CA). About 5 µL of the diluted digestion sample was injected with a mobile phase flow rate of 10 µL/min and gradient elution. Mobile phase A consisted of water with 0.1% formic acid, and mobile phase B consisted of acetonitrile with 0.1% formic acid. For human IgG, the HPLC gradient was as follows: 5% B for 3 min, 5% to 40% B in 37 min, 40% to 90% B in 10 min, 90% B for 10 min, 90% to 5% B in 10 min, and 5% B for 10 min. For the mixture of fetuin and avidin, the same solvents were used, and the HPLC gradient was: 2% B for 5 min, 2% to 45% B in 50 min, 45% to 90% B in 8 min, 90% B for 10 min, 90% to 2% B in 10 min, and 2% B for 10 min. A wash and blank run were applied between each sample to minimize sample carryover.

For mass spectrometric analysis, the positive ion mode was utilized with an ESI source voltage of 3 kV and capillary temperature of 250°C. The MS full scans were at a resolution of 30,000 (for *m/z* 400). The CID MS/MS scans were collected in a linear ion trap in a data-

dependent fashion. The ten most intense precursor ions from inclusion lists (when applicable) or from the MS preview scan were isolated for CID. The parent mass widths for inclusion lists were ±10 ppm. After being selected for MS/MS, with a repeat count of 2 within a repeat duration of 50 s, the precursor ion was dynamically excluded for 180 s. The FTMS had an automatic gain control (AGC) target value of $5e^5$ with a maximum injection time of 400 ms. For CID MS/MS, the AGC was set with a target value of 10^4 and a maximum injection time of 50 ms. The isolation mass window for selecting a precursor ion was 2 Da. Normalized collision energy of 30% was applied with an activation time of 10 ms. The MS and CID MS/MS data were manually interpreted to obtain the glycopeptide coverage.

3.2.4 Software Overview

GlycoPep MassList is a free publicly accessible software tool designed for generation of inclusion lists in targeted analysis of glycopeptides. The theoretical glycopeptide mass computation can be easily performed by specifying the protein sequence, glycan library, charge state, mass range, number of missed cleavages, and the isotope preferred by the users. The software was written in Java (JDK7) and can be run on Windows 7 or newer version of Windows. Java Runtime Environment 7 (JRE 7) is recommended to successfully run the software.

3.2.5 Glycan Library of the Software

To rapidly and effectively generate potential glycopeptide masses, a default glycan library, consisting of 340 glycans, was integrated into the software. These glycans include high mannose, hybrid, and complex glycans that are found in nature. The glycans were categorized into five groups: 1. high mannose and pauci-mannose glycans; 2. complex or hybrid glycans without sialic acids; 3. glycans with NeuAc; 4. glycans with NeuGc 5. glycans with [PO3] or [SO3]. For the sialylated glycans, those containing NeuAc were separated from the ones

containing NeuGc because NeuGc cannot be biosynthesized by humans [37]. Thus, when analyzing human derived glycoproteins, NeuGc-containing glycans are not present and do not need to be considered.

Although the native glycan library is extensive and contains biologically relevant glycans from a wide range of sources, it does not cover all the N glycans that are present in nature. Therefore, the software was designed so that users can upload their own glycan libraries. The users can also use one or multiple groups of glycans from the software together with their own glycan libraries.

3.2.6 Implementation of the Software for the Experiments Described Here

The glycopeptide *m/z*'s that were selected for inclusion list experiments in the examples shown herein were generated using these procedures: (1) No missed cleavages were calculated. (2) It was assumed that the glycopeptide did not have any post translational modification other than glycosylation. (3) The glycopeptide *m/z*'s from charge state +2 to +8 were considered. (4) The glycopeptide *m/z*'s in the range of *m/z* 800 to 2000 were selected. Procedures (1) and (2) were used to reduce the number of entries on inclusion list, yet keep the valuable entries. This increased the specificity of ion selection and decreased the probability of random matches between the masses on the list and the masses detected from the sample background that did not correspond to glycopeptide ions. While we have included the option to calculate glycopeptides with miscleavages, initial testing has shown that this option does not increase the number of unique glycopeptides subjected to MS/MS experiments, particularly when the protein is optimally digested initially. Procedure (3) was applied so that the glycopeptides with a variety of sizes could be identified using one search. The mass range (*m/z* 800 to 2000) was selected based on our experience on MS-based glycopeptide analysis: Almost all glycopeptides are observed in

at least one charge state above *m/z* 800. While these procedures may not be universal best practices, depending on the experiment to be performed, they demonstrated themselves to be advantageous in numerous test cases we ran. In addition to these procedures, we observed that it is worthwhile to select the precursor ion type carefully. The higher abundance ^{13}C mass of the glycopeptide was used instead of the lower abundance monoisotopic mass. In the experiments using human IgG, the first ^{13}C masses were used. In the experiments using a fetuin and avidin mixture, the second ^{13}C masses were used. These isotopes were selected based on the size of the peptide. The larger the peptide, the more likely that a higher isotope will be the most abundant peak in the isotopic cluster.

3.3 Results and Discussion

3.3.1 Software Development and Implementation

The software described herein is a free tool to calculate theoretical glycopeptide masses and generate inclusion lists for targeted data acquisition on glycopeptides. A screenshot of the user interface is shown in **Figure 3.1**. In order to calculate the glycopeptide mass, users input the required information such as protein sequence, glycan library, number of missed cleavages (up to 3), charge state (from 1 to 8) and mass range (400 to 2000). The users can also choose to calculate monoisotopic mass or the mass of one of the higher isotopes (up to the third ^{13}C mass), depending on the glycopeptide of interest.

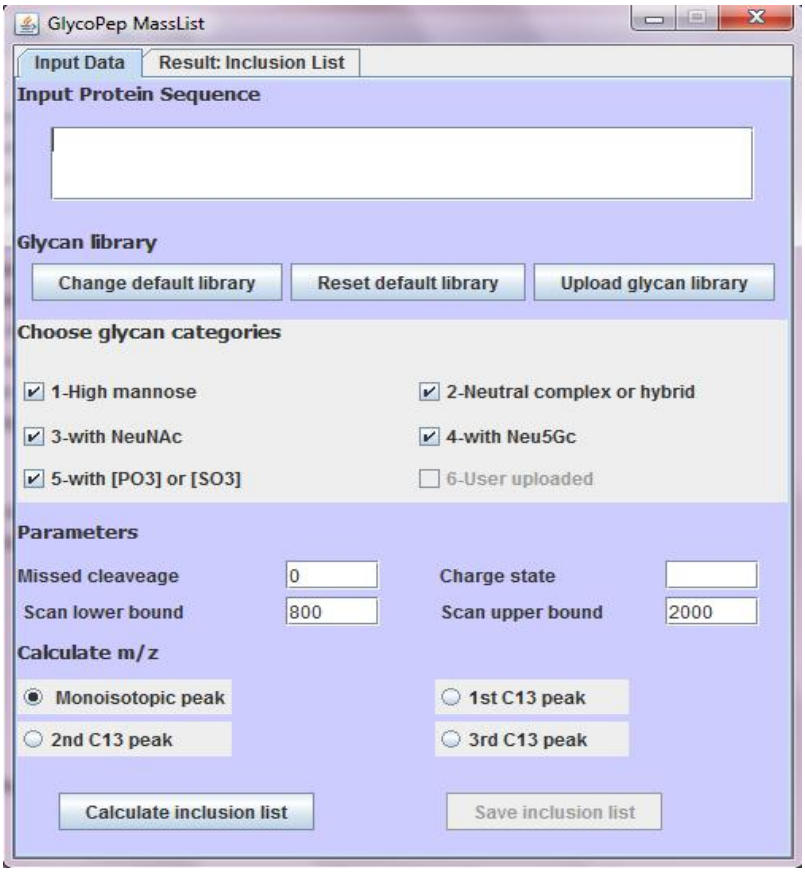


Figure 3.1 User interface for GlycoPep MassList. Users input the protein sequence, glycan library, charge state, mass range, number of missed cleavages, and select the mass (monoisotopic mass or ^{13}C mass) to be calculated. *In silico* tryptic digestion is then performed on the protein, and all the potential glycosylated peptides are reported by the software. The output is displayed under Result: Inclusion List. One or multiple groups of glycans can be selected, or custom glycan libraries can be uploaded, or a combination of manually input libraries and those in the database can be used.

After the appropriate data are input, the software executes an *in silico* tryptic digestion on the protein sequence, and all the peptides that have a potential N-linked glycosylation site (NXS/T, X≠P) are extracted. In the current form of the software, cysteines are, by default, considered to be alkylated by iodoacetamide (IAM), a widely used alkylation reagent in protein digestion. Next, theoretical glycopeptide *m/z* values, based on all the combinations of the glycosylated peptides and the user-specified glycans, are computed. The results are displayed under "Result: Inclusion List". Each theoretical glycopeptide appears on one line, with the glycopeptide composition shown on the left and *m/z* on the right.

After the software was developed and carefully tested, two sets of CID experiments were conducted, in order to compare the performance of targeted data acquisition strategy (with inclusion lists) and the conventionally used data dependent acquisition strategy (without inclusion lists), where the top 10 most intense precursor ions are selected for MS/ MS. The two experiments are henceforth referred to as the "inclusion" experiment, or the "top 10" experiment, respectively.

3.3.2 Analysis of Human IgG

Plasma-derived human IgG is an important class of antibody.[38-40] The glycosylation site located at the Fc region has an effect on the interaction of IgG with Fc gamma receptor (FcγR) [41]. Aberrant glycosylation on this site is related to diseases.[42] For example, a lower degree of galactosylation was observed in patients with rheumatoid arthritis.[8] When comparing the IgG glycosylation profiles of diseased and healthy states, one must identify all glycoforms present, otherwise the glycosylation differences, due to the samples themselves, will not be known. IgG is also a very important recombinantly expressed biotherapeutic.[43] The efficacy and side effect of IgG therapeutics are closely related to their glycosylation.[43,44] For instance, antibodies with high mannose glycans are cleared faster in human serum.[45] Glycans containing NeuGc can cause an unwanted immune response.[44] Thus, it is important to ensure the glycan profile fidelity across different expression vehicles, conditions, and batches.

Hence, human IgG was used to compare the efficiency of the new "inclusion List" software and strategy, to the traditional, untargeted, top 10 approach. Utilizing the software for the inclusion approach requires three steps: choosing an appropriate glycan library, building the inclusion list, and then conducting the LC-MS experiment. Each step is elaborated upon, briefly, here. When the software is used for other applications, a similar workflow should be followed.

An appropriate glycan list must be chosen. The N linked glycopeptides from IgG Fc region mainly contain complex-type biantennary glycans. These glycans are mostly core-fucosylated; they have up to two galactose residues; they may be sialylated and bisected by an N-acetylglucosamine (GlcNAc) residue.[46] We compiled a glycan library ideal for IgG analyses by researching the literature associated with human IgG glycosylation.[38-40,45,47] The glycan list for these experiments contains 52 glycans, and it can be found in the **SI Table 3.1**. All these glycans are either reported previously for IgG or are reasonable additional glycans that may be present, based on the rules of glycotransferase processing.

After building the glycan library, which can now be used for any IgG experiment, GlycoPep MassList was used to generate the inclusion masses for the targeted data acquisition. The IgG protein sequence and the compiled glycan library for human IgG were uploaded to obtain the theoretical glycopeptide masses. Since EEQYNSTYR and EEQFNSTFR glycopeptides are commonly observed in tryptic digests of IgG, they were both included. To compare the inclusion experiment with the commonly used top 10 experiment, LC-MS runs with an inclusion list and without the inclusion list were performed back-to-back on the same day using the same digested protein sample. In the inclusion experiment, an inclusion list populated with the first ^{13}C masses of the theoretical glycopeptides was imported into the instrument software, while in the top10 experiment, no inclusion list was applied.

The resulting MS data are depicted in **Figure 3.2**. The tryptic glycopeptides EEQYNSTYR and EEQFNSTFR eluted at about 6.5 min and 16.5 min, respectively. The elution of EEQFNSTFR is indicated by a pink bar on the total ion chromatogram (TIC) in **Figure 3.2a**. The MS data for this glycopeptide rich region of the chromatogram, containing multiple glycopeptides from the EEQFNSTFR site is shown in **Figure 3.2b**. The high abundance

glycopeptides in this spectrum, labeled with stars, were selected for CID in both the inclusion and top 10 experiments, while the low abundance ones, labeled with triangles, were only selected for CID in the inclusion experiment.

One glycopeptide identified only by inclusion but not top 10 is shown in **Figure 3.2c**. This low abundance glycopeptide was only observed in the +2 charge state in the high resolution MS data. Because of its relatively low abundance, this precursor ion was not selected for CID in the top 10 experiment. Nevertheless, in the inclusion experiment, this ion was selected for CID. As clearly shown in **Figure 3.2c**, the CID data still produced sufficient data to confirm the glycopeptide assignment.

All the detected glycopeptides, selected by inclusion and/or top 10, were identified. The data for each experiment are fully reported in **SI Table 3.2**. Most of the detected glycopeptides from each site were core fucosylated. A minor portion of them were sialylated. The sialylated glycopeptides eluted about one minute later than the non-sialylated ones, which agreed with previous reports.[39]

The glycopeptide coverage summary for the inclusion and top 10 experiments is shown in **Figure 3.3**. It is worth noting from Figure 3a that all the unique glycopeptides observed in the high resolution spectrum with ion intensities more than 200 counts (the signal threshold for CID) were selected for CID by the inclusion experiment. Additionally, inclusion outperformed top 10 by selecting seven more unique glycopeptides for CID, which contributed to a notable 39% higher glycopeptide coverage.

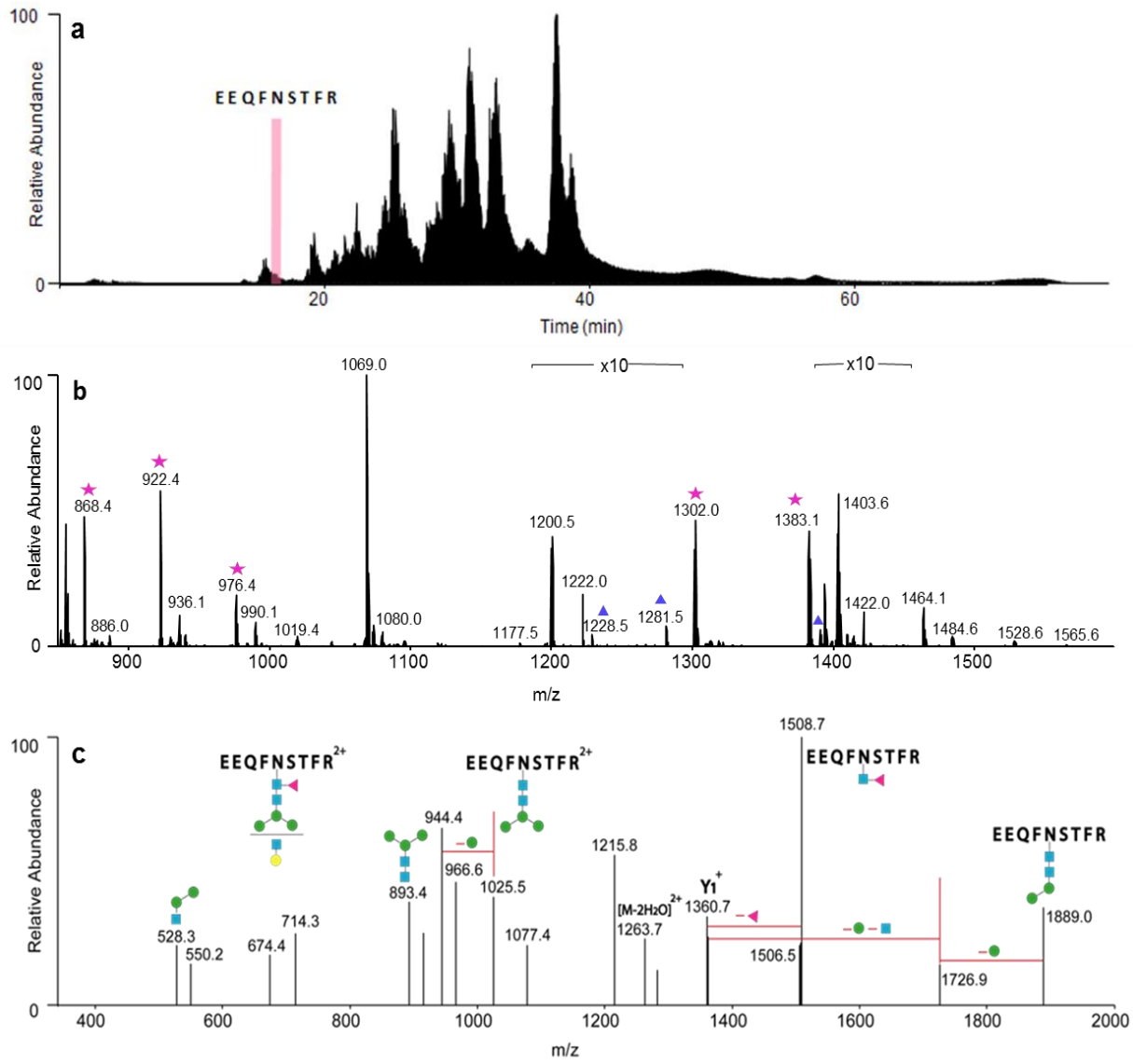


Figure 3.2 Example data from inclusion experiment. **a** Ion chromatogram of the IgG glycoprotein digest, indicating where one of the peptides, EEQFNSTFR, elutes. **b** MS data for the highlighted region in a. *Stars,* Glycopeptide selected for MS/MS by inclusion and top 10. *Triangles,* Glycopeptides selected for MS/MS only by inclusion but not by Top10. **c** The CID spectrum of a glycopeptide that was only selected by inclusion. The ion, *m/z* 1281.52, corresponds to EEQFNSTFR+[Hex]4[HexNAc]3Fuc]1. Its composition can be confirmed by the product ions in c. *Blue squares,* N-acetylhexosamines. *Red triangle,* fucose. *Green and yellow circles,* hexoses.

The number of unique glycopeptides from each site selected for CID by inclusion and top 10 are shown in **Figure 3.3b**. Inclusion showed higher efficiency than top 10 on both glycosylated peptides analyzed. The inclusion approach was especially advantageous on the site,

EEQFNSTFR. In this case, top 10 selected only 8 unique glycopeptides for MS/MS, while inclusion selected 13, showing a prominent advantage of 62% higher coverage. Although the advantage was narrower on site, EEQYNSTYR, where 12 unique glycopeptides were triggered CID by inclusion, and 10 by top 10, inclusion still showed 20% higher coverage.

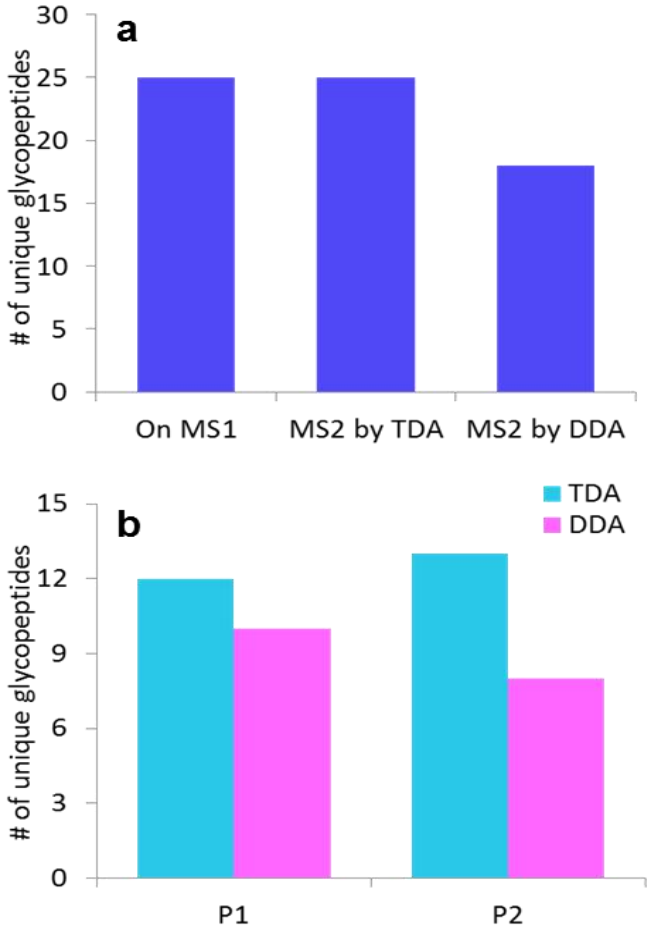


Figure 3.3 IgG glycopeptides selected for CID by inclusion or top 10 experiments. **a** Comparison of the total number of unique glycopeptides observed in the high resolution MS data vs. those that were selected for CID during inclusion or top 10. Inclusion got 100% coverage and selected seven more glycopeptides (39% more) than top 10. **b** The number of unique glycopeptides from each site selected for MS/MS by inclusion and top 10. *P1*=EEQYNSTYR. *P2*=EEQFNSTFR. For P2, a greater advantage was observed because the chromatogram had more co-eluting interferents in this region that were selected for MS/MS by the top 10 experiment

The reason the inclusion approach was more advantageous on one of these two sites is likely related to what else was co-eluting when the glycopeptides were being selected for CID.

The glycopeptide, EEQYNSTYR, eluted at about 6.5 min, while there was a limited number of non-glycosylated peptides co-eluting. The limited background from the non-glycosylated peptides allowed the traditional top 10 approach to pick up almost all of the glycopeptides for CID. The other glycopeptide, EEQFNSTFR, eluted at about 16.5 min, when the non-glycosylated peptides also started eluting. As the non-glycosylated peptides both outnumbered the glycopeptides, and were present in higher abundance, the efficiency of the traditional top10 experiment was substantially reduced: more duty cycle was wasted on the background peaks instead of glycopeptide peaks. On the contrary, the inclusion experiment was able to prioritize the duty cycle on the highly possible glycopeptide peaks pre-assigned on the inclusion list. This resulted in the inclusion approach working especially well for the EEQFNSTFR peptide, compared to EEQYNSTYR.

3.3.3 Targeted Data Analysis of Fetuin/Avidin Mixture

To further test the utility of the use of targeted inclusion lists for glycopeptide, we used a more complex sample involving two proteins with multiple glycosylation sites. More specifically, a mixture that had 15 pmol fetuin and 1.7 pmol avidin was used. This glycoprotein mixture was chosen for two reasons. First, these well studied glycoproteins have very different glycosylation profiles. Fetuin has complex type glycans, while avidin has mainly high mannose and hybrid type glycans.[48] Hence, the glycan library for this experiment incorporated all the three types of glycans. Secondly, the avidin protein was purposely added in a low abundance, since identifying low-abundant glycopeptides is a particularly large challenge in the field of glycoproteomics. We were interested in determining how the inclusion and top 10 strategies compared when the glycoprotein of interest was a small component of the overall sample.

In this experiment, the glycan library was designed differently from the first set of experiments as well. Unlike the case of human IgG, where most glycans of the glycan library are highly possible oligosaccharides for human IgG, most glycans in this library did not correspond to the ones in the glycoprotein sample used. Consequently, most *m/z*'s on the inclusion list did not correspond to the glycopeptide *m/z*'s in the sample. The goal of this set of experiments was to begin to understand the impact of using large inclusion lists; could these lists still select more glycopeptide ions for MS/MS than the commonly used top 10 approach?

The glycopeptide coverage summary for the inclusion and top 10 experiments is shown in **Figure 3.4**. **Figure 3.4a** summarizes the complete glycopeptide coverage for all the glycoforms at all the glycosylation sites, and, as expected, the inclusion approach was superior to the top 10 experiment in selecting more glycopeptide ions for CID. Similar to the results of human IgG, all the 27 unique glycopeptides observed in the high resolution mass data that had ion intensities above the signal threshold for CID were selected by inclusion for MS/MS experiments, while only 21 unique peptides were selected by top 10. This difference corresponds to a 29% increase in coverage.

To understand the circumstances that benefit the inclusion approach, we also compared the results for each protein separately. These are displayed in **Figure 3.4b**. In this figure, one can clearly see that the inclusion approach was most beneficial for the more complex, lower-abundant protein, avidin. For avidin, 16 unique glycopeptides were selected for MS/MS in inclusion, compared to 12 unique ones for top 10, representing a 33% higher coverage. For fetuin, inclusion was still superior, but the gain was smaller. Here, inclusion vs. top 10 resulted in 11 unique glycopeptides vs. 9 unique glycopeptides, a 22% higher coverage.

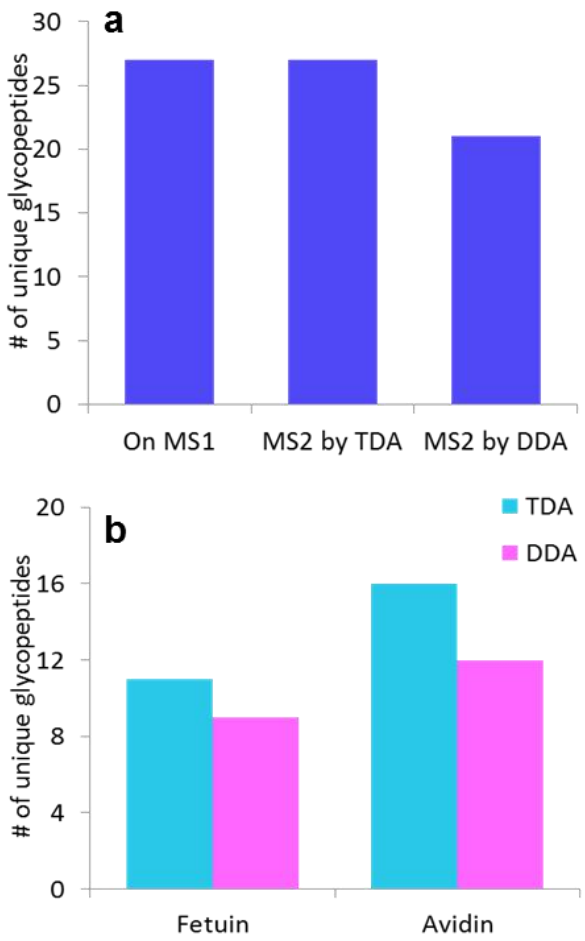


Figure 3.4 Result summary from the fetuin/avidin mixture. **a** All the unique glycopeptides observed in the high resolution spectrum were selected for MS/MS by inclusion. Inclusion had six more unique glycopeptides (29% more) selected for MS/MS than Top10. **b** The number of unique glycopeptides from each protein selected for MS/MS by inclusion and top 10. The inclusion approach was most efficacious for avidin because the top 10 experiments were less effective when many glycopeptides were present in a high background.

The results for the fetuin/avidin experiments are consistent with those from IgG. When glycopeptides are present in lower abundance, as is the case in the avidin experiment, or when other non-glycosylated peptides are co-eluting, as is the case in the IgG experiment, the inclusion approach shows its greatest benefit. These findings are not terribly surprising; in fact, we fully expected this to be the case because top 10 experiments are well-known to leave out MS/MS data on low-abundant species. What is somewhat surprising is that the inclusion approach worked much better, even on relatively simple samples. These findings suggest that inclusion approach

will show an even greater benefit when challenged with more complex samples, such as low-abundant glycopeptides present in high background of interfering ions.

For the experiments described here, the number of ions added to the inclusion list ranged between 233 ions for the IgG experiments and 1507 ions for the experiments where both fetuin and avidin were simultaneously examined. Clearly, the large number of ions on the inclusion list, compared with other more targeted approaches, did not limit the instrument's ability to pick all the relevant glycopeptides for MS/MS analysis. One key reason for the success of these experiments is the fact that we used a relatively small mass width (10 ppm) for the ions on the inclusion list. The mass width (Da) multiplied by the number of ions on the inclusion list roughly determines the amount of spectral space that is being queried in a given experiment. For example, with 1500 ions on the inclusion list, and a selection criterion of ±10 ppm, the spectral space being queried is approximately 0.03 Da multiplied by 1500 ions, or 45 Da, in the range of m/z 800–2000. In other words, about 4 % of the available spectral points between m/z 800 and 2000 are being queried under these circumstances. Either increasing the mass tolerance for the ions on the inclusion list (to larger than 10 ppm) or increasing the number of ions queried, increases the percentage of the spectral space being queried; therefore, these changes can decrease the value of using an inclusion list. The instrument control software on the mass spectrometer used for these studies limits the inclusion list to 2000 ions, so there is also a fixed limit to the number of ions that can be queried.

3.4 Conclusion

MS/MS data is necessary to accurately identify glycopeptides. Nonetheless, due to the relatively low signals of glycopeptides in MS scans, many glycopeptide ions are not selected for MS/MS during a single DDA experiment. This problem prompted us to develop the software

GlycoPep MassList to facilitate the inclusion experiments for glycopeptide analysis. This software tool can rapidly generate inclusion lists for glycoproteins, so that targeted glycopeptide analyses can be performed.

To test the application of this software and the inclusion strategy, two experiments were conducted. In both experiments, the inclusion strategy outperformed the traditional top 10 experiment by substantial margins. Furthermore, the experiments herein demonstrated that the inclusion approach is particularly advantageous when the glycoprotein of interest is present in low abundance, when it co-elutes with many non-glycosylated peptides, and/or when a variety of glycoforms are appended at the same glycosylation site. While the software was tested in several experiments where CID data were collected, the tool is agnostic towards the type of dissociation method used, and it could readily be applied to trigger ETD or HCD experiments as well.

The GlycoPep MassList software is freely available to any interested researchers.

3.5 Appendix-Supplemental Information

Table of Contents

SI table 3.1 Glycan library used in targeted analysis of human IgG Fc N-linked glycopeptides.

glycan composition	monoisotopic mass	glycan composition	monoisotopic mass
Hex]3[HexNAc]2	892.3172	[Hex]5[HexNAc]4[Fuc]1	1768.6395
[Hex]4[HexNAc]2	1054.3700	[Hex]4[HexNAc]5[Fuc]1	1809.6661
[Hex]3[HexNAc]3	1095.3966	[Hex]5[HexNAc]5	1825.6610
[Hex]5[HexNAc]2	1216.4229	[Hex]3[HexNAc]6[Fuc]1	1850.6926
[Hex]3[HexNAc]3[Fuc]1	1241.4545	[Hex]9[HexNAc]2	1864.6342
[Hex]4[HexNAc]3	1257.4494	[Hex]4[HexNAc]4[Fuc]1[NeuNAc]1	1897.6821
[Hex]3[HexNAc]4	1298.4760	[Hex]5[HexNAc]4[NeuNAc]1	1913.6770
[Hex]6[HexNAc]2	1378.4757	[Hex]4[HexNAc]5[NeuNAc]1	1954.7036
[Hex]4[HexNAc]3[Fuc]1	1403.5073	[Hex]5[HexNAc]5[Fuc]1	1971.7189
[Hex]5[HexNAc]3	1419.5022	[Hex]6[HexNAc]5	1987.7138
[Hex]3[HexNAc]4[Fuc]1	1444.5339	[Hex]5[HexNAc]4[Fuc]1[NeuNAc]1	2059.7349
[Hex]4[HexNAc]4	1460.5288	[Hex]4[HexNAc]5[Fuc]1[NeuNAc]1	2100.7615
[Hex]3[HexNAc]5	1501.5553	[Hex]5[HexNAc]5[NeuNAc]1	2116.7564
[Hex]7[HexNAc]2	1540.5285	[Hex]6[HexNAc]5[Fuc]1	2133.7717
[Hex]4[HexNAc]3[NeuNAc]1	1548.5448	[Hex]5[HexNAc]4[NeuNAc]2	2204.7724
[Hex]5[HexNAc]3[Fuc]1	1565.5601	[Hex]5[HexNAc]5[Fuc]1[NeuNAc]1	2262.8143
[Hex]6[HexNAc]3	1581.5551	[Hex]6[HexNAc]5[NeuNAc]1	2278.8092
[Hex]4[HexNAc]4[Fuc]1	1606.5867	[Hex]5[HexNAc]4[NeuNAc]2[Fuc]1	2350.8303
[Hex]5[HexNAc]4	1622.5816	[Hex]5[HexNAc]5[NeuNAc]2	2407.8518
[Hex]3[HexNAc]5[Fuc]1	1647.6132	[Hex]6 [HexNAc]5 [NeuNAc]1[Fuc]1	2424.8671

[Hex]4[HexNAc]5	1663.6082	[Hex]5[HexNAc]4[NeuNAc]3	2495.8678
[Hex]4[HexNAc]3[Fuc]1[NeuNAc]1	1694.6027	[Hex]5[HexNAc]5[NeuNAc]2[Fuc]1	2553.9097
[Hex]8[HexNAc]2	1702.5813	[Hex]6[HexNAc]5[NeuNAc]2	2569.9046
[Hex]3[HexNAc]6	1704.6347	[Hex]6[HexNAc]5[NeuNAc]2[Fuc]1	2715.9625
[Hex]6[HexNAc]3[Fuc]1	1727.6130	[Hex]6[HexNAc]5[NeuNAc]3	2861.0001
[Hex]4[HexNAc]4[NeuNAc]1	1751.6242	[Hex]6[HexNAc]5[NeuNAc]3[Fuc]1	3007.0580

SI table 3.2 All the detected glycopeptides from human IgG Fc region and whether each glycopeptide was selected for CID by TDA or/and DDA.

peptide	glycan	charge state	theoretical mass	experimental mass	mass error /ppm	MS2 by TDA?	MS2 by DDA?
EEQYNSTYR							
@	Hex3HexNAc3Fuc1	+2	1215.9869	1215.9844	2	yes	no
@	Hex3HexNAc4	+2	1244.4976	1244.4960	1	yes	no
		+3	830.0009	829.9995	2	yes	no
	Hex3HexNAc4Fuc1	+2	1317.5266	1317.5242	2	yes	yes
		+3	878.6868	878.6857	1	yes	yes
	Hex4HexNAc4	+2	1325.5240	1325.5221	1	yes	yes
		+3	884.0185	884.0164	2	yes	yes
	Hex4HexNAc4Fuc1	+2	1398.5530	1398.5501	2	yes	yes
		+3	932.7044	932.7034	1	yes	yes
	Hex5HexNAc4	+2	1406.5504	1406.5447	4	yes	no
		+3	938.0361	938.0347	1	yes	yes
	Hex3HexNAc5Fuc1	+2	1419.0663	1419.0647	1	yes	no
		+3	946.3799	946.3797	0	yes	yes
	Hex5HexNAc4Fuc1	+2	1479.5794	1479.5771	2	no	yes
		+3	986.7220	986.7215	1	yes	yes
	Hex4HexNAc5Fuc1	+2	1500.0927	1500.0895	2	no	no
		+3	1000.3976	1000.3976	0	yes	yes
	Hex4HexNAc4NeuNAc1Fuc1	+2	1544.1007	1544.0983	2	no	no
		+3	1029.7362	1029.7351	1	yes	yes
	Hex5HexNAc5Fuc1	+2	1581.1191	1581.1152	2	no	no
		+3	1054.4152	1054.4142	1	yes	no
	Hex5HexNAc4NeuNAc1Fuc1	+2	1625.1271	1625.1249	1	no	no
		+3	1083.7538	1083.7498	4	yes	yes
EEQFNSTFR							
	Hex3HexNAc3Fuc1	+2	1199.9920	1199.9906	1	yes	yes
@	Hex3HexNAc4	+2	1228.5027	1228.5028	0	yes	no
		+3	819.3376	819.3318	7	no	no
@	Hex4HexNAc3Fuc1	+2	1281.0184	1281.0170	1	yes	no
	Hex3HexNAc4Fuc1	+2	1301.5317	1301.5293	2	yes	yes
		+3	868.0236	868.0219	2	yes	yes
@	Hex4HexNAc4	+2	1309.5291	1309.5249	3	yes	no

		+3	873.3552	873.3538	2	no	no
	Hex4HexNAc4Fuc1	+2	1382.5581	1382.5548	2	yes	yes
		+3	922.0412	922.0398	2	yes	yes
@	Hex5HexNAc4	+2	1390.5556	1390.5475	6	yes	no
		+3	927.3728	927.3713	2	no	no
	Hex3HexNAc5Fuc1	+2	1403.0714	1403.0701	1	yes	yes
		+3	935.7167	935.7162	1	yes	no
	Hex5HexNAc4Fuc1	+2	1463.5845	1463.5808	3	no	no
		+3	976.0588	976.0574	1	yes	yes
	Hex4HexNAc5Fuc1	+2	1484.0978	1484.0966	1	no	no
		+3	989.7343	989.7338	1	yes	yes
	Hex4HexNAc4NeuNAc1Fuc1	+2	1528.1058	1528.1031	2	no	no
		+3	1019.0730	1019.0726	0	yes	yes
@	Hex5HexNAc5Fuc1	+2	1565.1242	1565.1205	2	no	no
		+3	1043.7519	1043.7510	1	yes	no
	Hex5HexNAc4NeuNAc1Fuc1	+2	1609.1322	1609.1289	2	no	no
		+3	1073.0906	1073.0894	1	yes	yes

@ indicates the unique glycopeptide that was selected for CID by TDA but not by DDA.

3.6 References

(1) Dwek, R. A. *Chem Rev* **1996**, *96*, 683-720.

(2) Deribe, Y. L.; Pawson, T.; Dikic, I. *Nat Struct Mol Biol* **2010**, *17*, 666-672.

(3) Raman, R.; Raguram, S.; Venkataraman, G.; Paulson, J. C.; Sasisekharan, R. *Nat Methods* **2005**, *2*, 817-824.

(4) Harvey, D. J. *Expert Rev Proteomics* **2005**, *2*, 87-101.

(5) Dell, A.; Morris, H. R. *Science* **2001**, *291*, 2351-2356.

(6) Go, E. P.; Rebecchi, K. R.; Dalpathado, D. S.; Bandu, M. L.; Zhang, Y.; Desaire, H. *Anal Chem* **2007**, *79*, 1708-1713.

(7) Wormald, M. R.; Dwek, R. A. *Structure* **1999**, *7*, R155-R160.

(8) Nagae, M.; Yamaguchi, Y. *Int. J. Mol. Sci.* **2012**, *13*, 8398-8429.

(9) Kolarich, D.; Lepenies, B.; Seeberger, P. H. *Curr Opin Chem Biol* **2012**, *16*, 214-220.

(10) Leymarie, N.; Griffin, P. J.; Jonscher, K.; Kolarich, D.; Orlando, R.; McComb, M.; Zaia, J.; Aguilan, J.; Alley, W. R.; Altmann, F.; Ball, L. E.; Basumallick, L.; Bazemore-Walker, C. R.; Behnken, H.; Blank, M. A.; Brown, K. J.; Bunz, S.-C.; Cairo, C. W.; Cipollo, J. F.; Daneshfar, R., et al. *Molecular & Cellular Proteomics* **2013**, *12*, 2935-2951.

(11) Grass, J.; Pabst, M.; Chang, M.; Wozny, M.; Altmann, F. *Anal Bioanal Chem* **2011**, *400*, 2427-2438.

(12) Fuster, M. M.; Esko, J. D. *Nature Reviews Cancer* **2005**, *5*, 526-542.

(13) Gornik, O.; Lauc, G. *Disease Markers* **2008**, *25*, 267-278.

(14) Saroha, A.; Biswas, S.; Chatterjee, B. P.; Das, H. R. *Journal of chromatography. B, Analytical technologies in the biomedical and life sciences* **2011**, *879*, 1839-1843.

(15) Perdivara, I.; Deterding, L. J.; Cozma, C.; Tomer, K. B.; Przybylski, M. *Glycobiology* **2009**, *19*, 958-970.

(16) Fujimura, T.; Shinohara, Y.; Tissot, B.; Pang, P. C.; Kurogochi, M.; Saito, S.; Arai, Y.; Sadilek, M.; Murayama, K.; Dell, A.; Nishimura, S.; Hakomori, S. I. *Int J Cancer* **2008**, *122*, 39-49.

(17) Qiu, Y.; Patwa, T. H.; Xu, L.; Shedden, K.; Misek, D. E.; Tuck, M.; Jin, G.; Ruffin, M. T.; Turgeon, D. K.; Synal, S.; Bresalier, R.; Marcon, N.; Brenner, D. E.; Lubman, D. M. *Journal of proteome research* **2008**, *7*, 1693-1703.

(18) Drake, P. M.; Schilling, B.; Niles, R. K.; Prakobphol, A.; Li, B.; Jung, K.; Cho, W.; Braten, M.; Inerowicz, H. D.; Williams, K.; Albertolle, M.; Held, J. M.; Iacovides, D.; Sorensen, D. J.; Griffith, O. L.; Johansen, E.; Zawadzka, A. M.; Cusack, M. P.; Allen, S.; Gormley, M., et al. *J. Proteome Res.* **2012**, *11*, 2508-2520.

(19) Stowell, S. R.; Ju, T.; Cummings, R. D. *Annual Review of Pathology* **2015**, *10*, 473-510.

(20) Ohtsubo, K.; Marth, J. D. *Cell* **2006**, *126*, 855-867.

(21) Dube, D. H.; Bertozzi, C. R. *Nature Reviews Drug Discovery* **2005**, *4*, 477-488.

(22) Goldberg, D.; Bern, M.; Parry, S.; Sutton-Smith, M.; Panico, M.; Morris, H. R.; Dell, A. *Journal of proteome research* **2007**, *6*, 3995-4005.

(23) Desaire, H. *Molecular & Cellular Proteomics* **2013**, *12*, 893-901.

(24) Dalpathado, D. S.; Desaire, H. *Analyst* **2008**, *133*, 731-738.

(25) Desaire, H.; Hua, D. *International Journal of Mass Spectrometry* **2009**, *287*, 21-26.

(26) Mechref, Y. *Current protocols in protein science / editorial board, John E. Coligan ... [et al.]* **2012**, *12*, 1-11.

(27) Quan, L.; Liu, M. *Modern Chemistry & Applications,* **2012**, *1*, e102.

(28) Aldredge, D.; An, H. J.; Tang, N.; Waddell, K.; Lebrilla, C. B. *J. Proteome Res.* **2012**, *11*, 1958-1968.

(29) Li, F.; Glinskii, O. V.; Glinsky, V. V. *Proteomics* **2013**, *13*, 341-354.

(30) Gagneux, P.; Varki, A. *Glycobiology* **1999**, *9*, 747-755.

(31) Schmidt, A.; Gehlenborg, N.; Bodenmiller, B.; Mueller, L. N.; Campbell, D.; Mueller, M.; Aebersold, R.; Domon, B. *Molecular & Cellular Proteomics* **2008**, *7*, 2138-2150.

(32) Liu, T.; Qian, W.-J.; Gritsenko, M. A.; II, D. G. C.; Monroe, M. E.; Moore, R. J.; Smith, R. D. *Journal of proteome research* **2005**, *4*, 2070-2080.

(33) Madera, M.; Mechref, Y.; Novotny, M. V. *Analytical Chemistry* **2005**, *77*, 4081-4090.

(34) Hart-Smith, G.; Low, J. K.; Erce, M. A.; Wilkins, M. R. *J. Am. Soc. Mass Spectrom.* **2012**, *23*, 1376-1389.

(35) Wu, Y.; Mechref, Y.; Klouckova, I.; Mayampurath, A.; Novotny, M. V.; Tang, H. *Rapid Communications in Mass Spectrometry* **2010**, *24*, 965-972.

(36) Froehlich, J. W.; Dodds, E. D.; Wilhelm, M.; Serang, O.; Steen, J. A.; Lee, R. S. *Molecular & Cellular Proteomics* **2013**, mcp. M112. 025494.

(37) Varki, A. *American journal of physical anthropology* **2001**, *116*, 54-69.

(38) Stadlmann, J.; Pabst, M.; Kolarich, D.; Kunert, R.; Altmann, F. *Proteomics* **2008**, *8*, 2858-2871.

(39) Selman, M. H. J.; Derks, R. J. E.; Bondt, A.; Palmblad, M.; Schoenmaker, B.; Koeleman, C. A. M.; van de Geijn, F. E.; Dolhain, R. J. E. M.; Deelder, A. M.; Wuhrer, M. *Journal of Proteomics* **2012**, *75*, 1318-1329.

(40) Zauner, G.; Selman, M. H. J.; Bondt, A.; Rombouts, Y.; Blank, D.; Deelder, A. M.; Wuhrer, M. *Molecular & cellular proteomics : MCP* **2013**, *12*, 856–865.

(41) Vidarsson, G.; Dekkers, G.; Rispens, T. *Front Immunol* **2014**, *5*, 1-17.

(42) Shade, K.-T.; Anthony, R. *Antibodies* **2013**, *2*, 392-414.

(43) Jefferis, R. *Nature reviews. Drug discovery* **2009**, *8*, 226-234.

(44) Ghaderi, D.; Taylor, R. E.; Padler-Karavani, V.; Diaz, S.; Varki, A. *Nature biotechnology* **2010**, *28*, 863-867.

(45) Shah, B.; Jiang, X. G.; Chen, L.; Zhang, Z. *J. Am. Soc. Mass Spectrom.* **2014**, *25*, 999-1011.

(46) Reusch, D.; Haberger, M.; Maier, B.; Maier, M.; Kloseck, R.; Zimmermann, B.; Hook, M.; Szabo, Z.; Tep, S.; Wegstein, J.; Alt, N.; Bulau, P.; Wuhrer, M. *mAbs* **2015**, *7*, 167-179.

(47) Wuhrer, M.; Stam, J. C.; van de Geijn, F. E.; Koeleman, C. A.; Verrips, C. T.; Dolhain, R. J.; Hokke, C. H.; Deelder, A. M. *Proteomics* **2007**, *7*, 4070-4081.

(48) Oliver, R. W. A.; Greent, B. N.; Harvey, D. J. *Biochem Soc Trans* **1996**, *24*, 917-927.

CHAPTER IV

Characterizing Spectral Flow Cytometry with Time-Delayed Integration (TDI SFC)

4.1 Introduction

Multi-parameter flow cytometry (MFC) is routinely used to characterize biological cells for size, granularity, and protein expression. MFCs typically consist of a sheath flow system to hydrodynamically align immunolabeled cells into a single file through an incident laser beam(s) to provide high sampling efficiency and reduced variation in the signal due to the Gaussian intensity profile of the excitation beam. Optical bandpass filters spectrally sort fluorescence into relatively wide wavelength ranges (channels) that are transduced by photomultiplier tubes (PMTs). Modern MFCs analyze 10^4 –10^5 cells/s on different fluorescence channels, enabling the ability to process cells and identify unique subpopulations.[1] However, there are intrinsic limitations associated with processing low numbers of cells (<1,000) using MFC, such as the need for establishing proper gating thresholds and compensation for channel cross-talk requiring high (>20,000) cells used for the analysis. Further, with poor spectral resolution, which is determined by the filter pass band, there is an inherent risk that biological "noise", such as abnormal autofluorescence from apoptotic cells,[2,3] can produce misclassification.[1] Improvements in immunostaining classification by morphological localization of fluorescence signals, typically restricted to microscopy, has been demonstrated using imaging flow cytometers (IFCs), such as the ImageStream® that can acquire high-resolution, multispectral microscopy images of flowing cells.[4-8] Yet such systems are costly, and for routine clinical analysis, throughput is limited by semi-automated data analysis.

Spectral flow cytometry (SFC) shares the same hydrodynamic sheathing and laser excitation strategy as MFC, but fluorescence emission is spectrally sorted onto a multichannel detector, yielding higher spectral resolution than MFC. The emission spectrum can be deconvoluted to quantify multiple fluorophores with considerable spectral overlap, including autofluorescence signals.[3,9-13] A commercial SFC instrument from Sony uses a series of prisms to spectrally disperse fluorescence onto a 32-channel PMT array (500-800 nm, 9.4 nm resolution).[3,11] The Nolan lab pioneered the use of a diffraction grating spectrograph to disperse fluorescence onto a CCD camera with high spectral resolution (0.4 nm/pixel).[13] However, the advantages provided by the spectrograph's resolution are countered by the operating characteristics of the CCD.[10,12,13] The sensor's duty cycle, the percent of time the sensor is actively imaging, is low (~0.1%) due to short exposure times (~10 μs) and the extended readout times (~10 ms).[13] Thus, the probability that cells would not be detected is high, which can be problematic for applications requiring the analysis of rare, low abundant cells.[14]

We present an SFC system operating in a time-delayed integration mode (TDI SFC).[8,15] The optical schematic and photograph of the TDI SFC is shown in **Figure 4.1**. Cells are sheathed into the cytometer's field-of-view, where 488 nm laser excitation is coincident with the observation zone of the collection optics. A spectrograph spectrally sorts the fluorescence along the columns of the CCD pixel array, which is spatially confined to a few rows of the CCD (see **Figure 4.2**). As the cell moves through the field-of-view, the image is shifted towards the serial register at a rate that corresponds to the cell's linear velocity (*i.e.*, synchronization). As the cell exits the field-of-view, the integrated emission spectrum is delivered for readout to the serial register. As such, the shutter of the CCD sensor is rarely shut with the duty cycle determined by the readout time of the CCD's serial register.

The coupling of SFC with TDI provides two main advantages. (i) Integration of the signal in TDI mode along n pixel columns of the CCD improves the signal-to-noise by $\sqrt{n}$ compared to a full frame readout in traditional SFC, which involves a snapshot mode of operation.[15] (ii) The shutter rarely closes (duty cycle approaches 100%) and multiple cells occupying the field-of-view can be resolved by the short readout time. Thus, TDI SFC is amenable for automated immunophenotyping of rare cells when the number of total cells available is limited.

The TDI SFC system consisted of a microfluidic device providing 1D sheathing that was interfaced with a custom optical system possessing 488 nm laser excitation, a collection microscope objective, long pass filters to remove the Rayleigh scattered radiation and the laser light, spectrograph, and a CCD sensor operated in a TDI mode. In this chapter, we optimized the operation of this TDI SFC in terms of synchronization of the CCD TDI readout with the flow velocity of biological cell models, in this case fluorescent polystyrene microbeads. The acquired data was processed in MATLAB and statistically analyzed to discover the flow rates that gave the best synchronization. Subsequently, we performed additional optimizations to improve upon the SNR. We also investigated the dynamic range of the TDI SFC to discover the upper throughput limit. All of these characterizations were critical and necessary to facilitate successful follow-up experiments on real biological samples, such as circulating leukemia cells (CLCs) isolated from the peripheral blood of leukemia patients or circulating tumor cells from epithelial cancer patients.

4.2 Experimental Methods

4.2.1 Reagents and Materials

Microfluidic devices were fabricated using polymethylmethacrylate (PMMA) substrates

(Plaskolite, Inc.) and PEEK tubing (0.020" ID, 1/32" OD and 0.007" ID, 1/32" OD, Idex Health and Science) that was used to connect the input port of the microfluidic to a syringe pump. Custom capillary connectors were Inner-Lok™ union capillary connectors (Polymicro Technologies) and barbed socket Luer Lock™ fittings (3/32" ID, McMaster-Carr). Reagents included: methanol-stabilized formaldehyde (37%), methanol-free formaldehyde (16%), phosphate buffered saline (1X PBS, pH=7.4), and InSpeck™ Green fluorescent calibration kit from Thermo Fisher Scientific; sucrose from Fluka; Micro-90®, reagent-grade isopropanol (IPA) from Sigma-Aldrich; and anti-mouse Fc-specific Antibody Binding Capacity (ABC) beads from Bangs Lab.

4.2.2 TDI SFC Optics

A 488 nm solid-state laser (54 mW, TEM_{00}, $1/e^2$=1.3 mm, Vortran Laser Technology) was used as the excitation source with the beam expanded using a 21.2× Keplerian beam expander (planoconvex lenses, f=3 mm, f=63.5 mm, VIS 0° coated; Edmund Optics) or a 7.1× Galilean beam expander (biconcave lens, f=-9 mm and planoconvex lens, f=63.5 mm, -A and VIS 0° coated; Thorlabs and Edmund Optics, respectively). The beam was filtered using a 488/6 nm BrightLine® bandpass filter (Semrock), and the wings of the beam were cut using an adjustable diaphragm (4.7 mm; Thorlabs) and focused (planoconvex lens, f=105 mm, -A coated; Thorlabs) to the back focal plane of a microscope objective (10x Fluar, NA=0.5, WD=1.9 mm or 20x Fluar, NA=0.75, WD=0.6 mm; Zeiss). Emission was collected via epi-illumination, which was filtered through a 495 nm BrightLine® dichroic beamsplitter and a 496/LP BrightLine® edge filter (Semrock) and diverted to a CP200 spectrograph (Jovin Yvon) using silver turning mirrors (Thorlabs). Fluorescence spectra were transduced using a Spec 10-100B back-illuminated CCD (1340×100 pixels, 20 μm, Princeton Instruments; 16-bit dynamic range)

controlled via a custom program built using the Scientific Instrument Tool Kit for LabVIEW® (R Cubed Software). The system contained a focusing camera (MU503 RGB CMOS camera, AmScope) inserted into the optical train via a kinematic beam turning cage cube and tube lens (f=165 mm, -A coated; Thorlabs). All optics were assembled and mounted using cage cubes, lens tubes, and micrometer translational stages (Thorlabs) and were enclosed in a light tight box. Power measurements were made with a S170C sensor (Thorlabs).

4.2.3 TDI SFC Data Processing

Acquired spectral data was streamed from LabVIEW to a TDMS file. A custom, off-line MATLAB code converted the TDMS data, reconstructed the spectral frames, and performed subtraction of a smoothing spline background fit to 2000 consecutive frames not containing signal. Frames without any event present were eliminated based on an S/N threshold of 20 (fluorescent beads) or 10 (cell samples). The resulting positive frames were analyzed to extract event properties regarding time duration and fluorescence intensity. The detailed data processing workflow can be found in **SI Figure 4.1**.

4.2.4 Microfluidic flow cell fabrication and fluidic connections

A brass mold master for the microfluidic flow cell was prepared using high precision-micromilling (KERN 44, KERN Micro- und Feinwerktechnik) and carbide bits (Performance Micro Tool). Devices were fabricated by hot embossing; upper platen temperature was 155 °C; lower platen temperature was 80 °C; pressure used for embossing was 1250 psi with a 3 min molding time, and demolding at 102 °C. A mold release agent (MoldWiz® F57-NC) was kindly gifted from Axel Plastics and was coated on the molding tool. Embossed devices were diced with a bandsaw, mechanically drilled to make ports, cleaned with 10% Micro-90®, IPA, and nanopure water before thermal fusion bonding (105 °C, 2 h) with cleaned coverslips (250 μm)

made from the same material as the chip substrate. PEEK tubing was epoxy sealed to the input/output ports. Microchannel dimensions were measured before and after thermal fusion bonding using a laser-scanning, confocal profilometer (VK-X250, Keyence). The detailed microfluidic flow cell fabrication and characterization information can be found in **SI Figure 4.2**. The flow cell was mounted to the TDI SFC system via a 3-axis micrometer translational stage (Thorlabs) and connected to two PicoPlus syringe pumps (Harvard Apparatus), one driving the sample flow and another driving both sheath flows. Sample was introduced via a 2-position, 6-port, manual injection valve (Idex Health and Scientific) with a PEEK, 20 µL injection loop.

4.2.5 TDI SFC System Optimization

For data without flow focusing, we infused InSpeck™ Green fluorescent calibration beads (100% intensity) at 0.675 µL/min (average 2 mm/s) into a square glass capillary (75 µm ID; Polymicro Technologies), that was mounted into the TDI SFC system and a section burned to expose the imaged field-of-view. To synchronize flow velocity with TDI clocking, we infused 100% intensity calibration beads (~800 beads/µL) in 15% sucrose/PBS. We used dimmer (1% intensity beads) for the collection of data to allow for the comparison of 10X and 20X microscope objectives.

4.2.6 Pixel Calibration

To calibrate the pixels on the spectral direction to wavelength values, we used an optical fiber to conduct the light source at different wavelengths with a 5 nm range at each wavelength from a fluorometer and introduced the light into our TDI SFC. The data our TDI SFC recorded at each wavelength were acquired and processed in MATLAB. The resulting data gave the pixel number that showed the highest signal at each wavelength, allowing for pixel calibration.

4.2.7 Simulations

Linear velocity profiles in the flow cell after merging of the sample and sheath flow streams were simulated (COMSOL Multiphysics 5.2a) using geometry dimensions obtained from laser scanning profilometry (**SI Table 3**). We simulated TDI readout with a custom time-stepping simulation coded in Fortran 90 and compiled with GFortran (MinGW).

4.2.8 Statistical Analysis

Data sets were tested for normality via Shapiro-Wilk Normality Testing. We compared non-normal data sets by Wilcoxon Rank Sum and Signed Rank Testing with R Studio v1.0.153 and R v3.5.1. Reported medians include all data. Box plot outliers were outside upper/lower quartile ± interquartile range.

4.3 Results and Discussion

4.3.1 Optical Design of the TDI SFC System

The optical schematic and photograph of the TDI SFC is shown in **Figure 4.1**. For excitation (shown in blue), we expanded a 488 nm laser (54 mW, spectrally filtered 488/6 nm) and focused the beam to the back focal plane of a high numerical aperture, infinity-corrected microscope objective (either 10X, 0.50 NA or 20 X, 0.75 NA) to establish widefield epi-illumination. We tested two beam expanders to enlarge the laser excitation beam (21.2× or 7.1×). In each case, we used a circular aperture to eliminate light outside the sensor's visible field-of-view (Köhler epi-illumination), which also improved excitation uniformity by cutting out the edges of the laser's Gaussian profile. With the 21.2× beam expander, we measured 4 mW (10.7 mW/cm2) excitation light through the microscope objective. Consolidating excitation with the 7.1× expander increased power to 21 mW (56.1 mW/cm2) but with less uniform flux (RSD in power density increased from 2% to 15 %,).

We collected fluorescence emission (shown in yellow in **Figure 4.1A**) through the same microscope objective and split it from the excitation path via a dichroic mirror (495 LP). We filtered emission (496 LP) before using an f-matched lens to focus the image onto a spectrograph's diffraction grating, which splayed the spectrally sorted emission along one axis of a back-illuminated CCD. In our optical systems there was a removable mirror and two cameras, one for TDI detection and one for system alignment. Because the TDI CCD did not provide full-fame images when collecting spectral data from the illuminated specimen, it was difficult for operators to know if the specimen was well aligned with respect to the imaging microscope objective. Thus, we included a kinematically removable mirror to bypass the TDI SFC emission optics and reconstruct the image on a standard CMOS camera via a tube lens. After alignment, the mirror was removed to enable TDI SFC imaging.

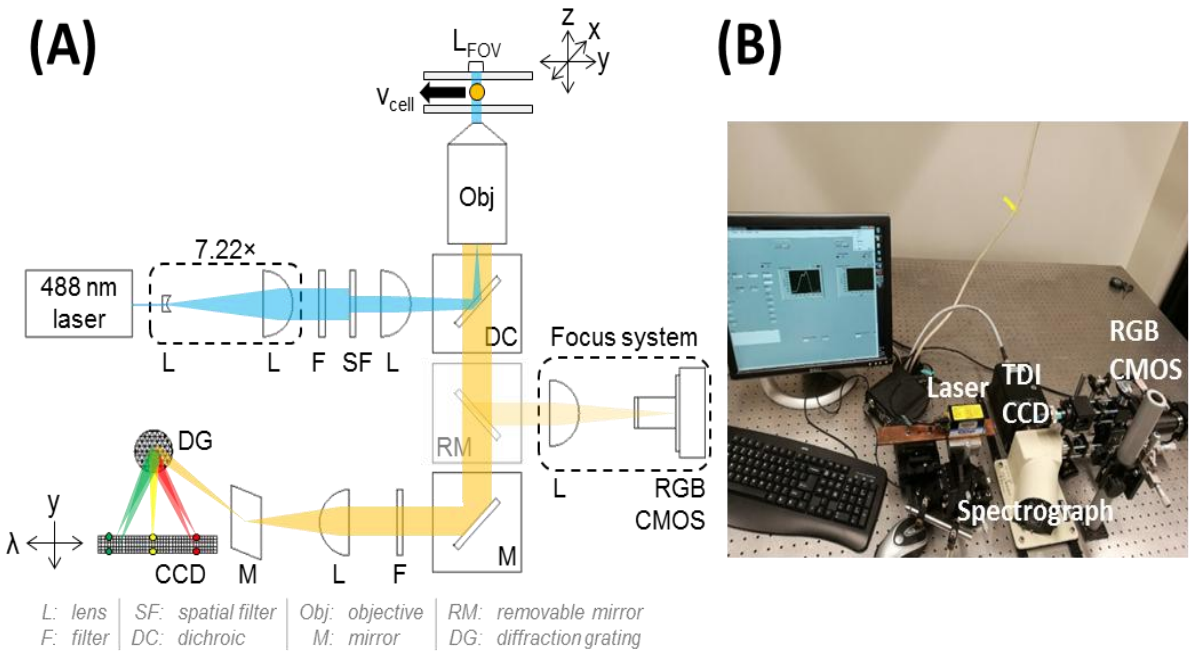


Figure 4.1 (A) Schematic diagram of the TDI spectral flow cytometer. A beam expander and epi-illumination lens produce widefield excitation in the flow cell from the filtered 488 nm laser. Fluorescence emission is filtered and passed to a spectrograph grating that spectrally sorts and focuses emission onto a back-illuminated CCD. A RGB CMOS camera was installed to serve as an alignment camera. (B) A photograph of our in-house developed TDI spectral flow cytometer.

4.3.2 Principle of TDI Synchronization

A schematic showing the operation of the TDI SFC is depicted in **Figure 4.2A**. The key to detecting a fluorescently labeled cell using time-delayed integration resides in two CCD operating aspects: 1) The cell's image moves in the same direction as the CCD's parallel shift direction; and 2) the cell's linear velocity should be synchronized with the CCD's readout rate. When these two conditions are met, the photoelectrical signal accumulates in only a few pixel rows at readout, increasing the signal to noise ratio (SNR) in the resultant image [8]. If the particles' linear velocity and the CCD's readout rate are not synchronized well, there will be image "slur" [9], and the photoelectrical signal will be distributed in multiple pixel rows, reducing the resultant SNR in the particle. With TDI, the CCD recorded a spectrum on each pixel row.

To facilitate TDI operation, a CCD detector array with 100 pixel rows by 1340 pixel columns was installed within the TDI SFC. The 100 pixels corresponded to the shift axis or time axis while the 1340 pixels corresponded to the spectral axis. On the spectral axis, each pixel corresponded to a wavelength value. The pixel calibration curve that converted pixel number to wavelength is shown in **Figure 4.2B**. As illustrated in **Figure 4.2B**, our system was capable of a spectral resolution of 0.7 nm/pixel with a fairly wide spectral range (500-1000 nm). The rectangle shape of the CCD made sure that the spectral axis was wide enough to encompass sufficient wavelength range. For the shift or time axis, 100 pixel rows corresponded to a 10-fold increase in the SNR compared to a snapshot mode of CCD operation.

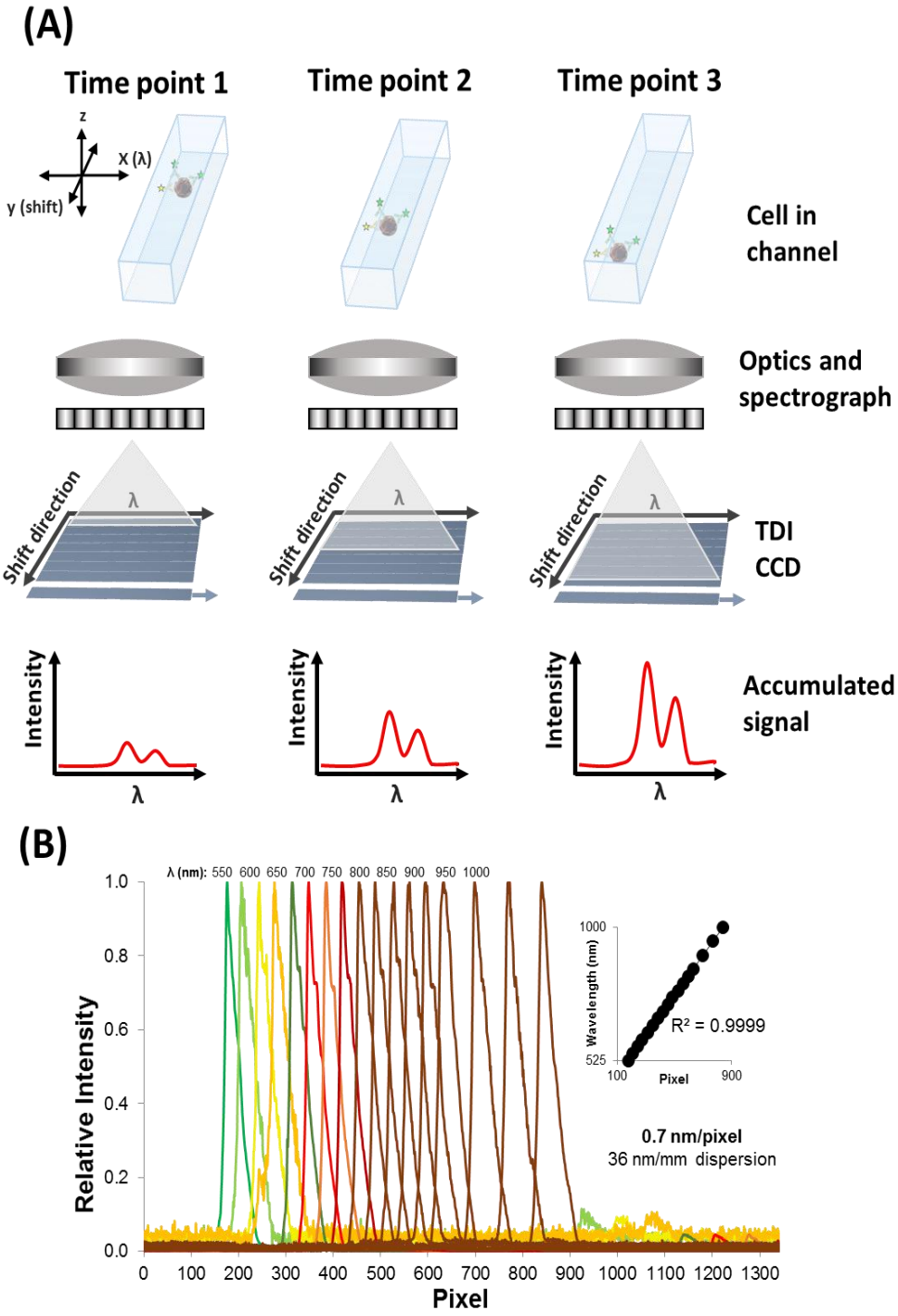


Figure 4.2 (A) Schematic detecting a fluorophore-labeled cell in our TDI SFC. Because the cell's linear velocity and the CCD's readout rate are synchronized, as the cell moving forward on the shift or time axis, the fluorescent signal is transported and integrated within only a few pixel rows at readout. Each pixel row on spectral direction collects a full spectrum due to the spectrum dispersion by the spectrograph. (B) Pixel calibration on the spectral direction. Each pixel on spectral direction corresponds to a specific wavelength value. Our system provides a high spectral resolution of 0.7 nm/pixel and a wide spectral range (500-1000 nm) for data acquisition.

If perfectly synchronized, the cell's velocity will match the clocking of the fluorescence spectrum, and the collected spectrum will be confined to a few readouts. For example, if the image resolution of the system is 6 μm/pixel, a perfectly tracked 6 μm cell will register in only one readout; with image size of 2 μm/pixel, three readouts. However, if a cell moves faster or slower than the CCD's time delay, the collected signal will "slur" over more readouts. We show this effect theoretically in **Figure 4.3**. Mismatching between velocity and TDI shifting can cause the CCD to readout fluorescence over longer time durations. Rather than a peak in signal, the signal decreases and broadens, risking that the spectrum will not be detected above background noise.

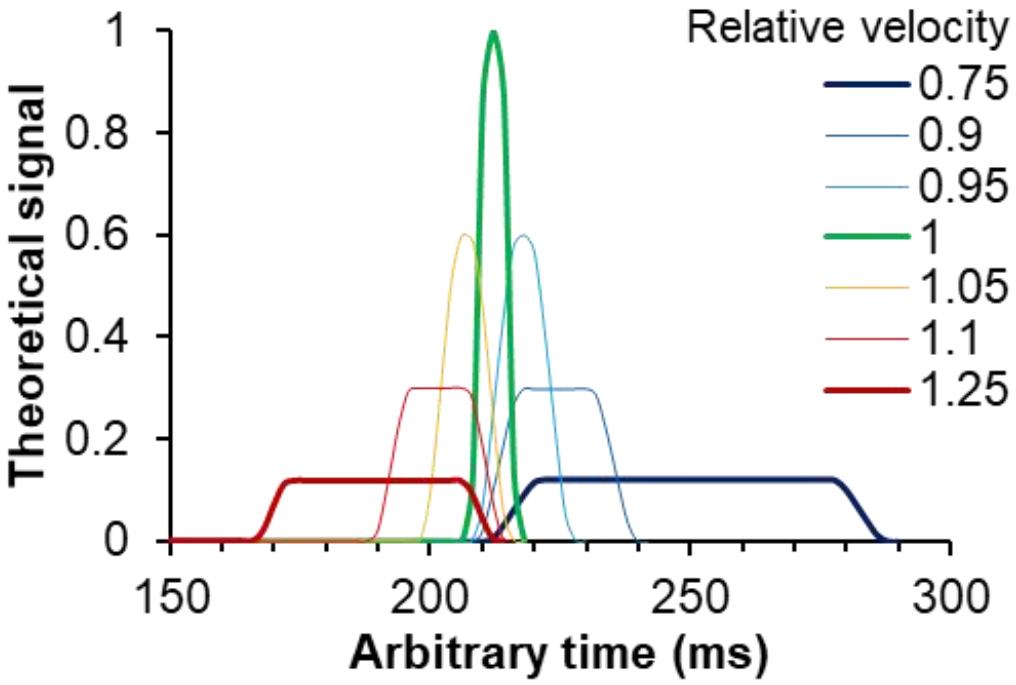


Figure 4.3 Theoretical TDI peak shapes for varying synchronization between cell velocity and CCD delay time, where relative velocity of 1 indicates perfect synchronization.

To initially gauge the necessary velocity, we measured the readout time for the analog-to-digital conversion of a single spectrum using our CCD, which was a minimum t_{delay} of 2 ms. The measurement of the minimum delay time per pixel row is detailed in SI (**SI table 4.1**). Over the

CCD's 100 rows, the minimum integration time was then 200 ms. We measured the field-of-view (FOV) of the TDI-CCD with a 10X objective (393 μm, 3.93 μm/pixel) by reference to the focusing CMOS camera. The detailed measurements are given in SI (**SI Figure 4.3** and **SI table 4.2**). Thus, for optimal TDI synchronization, the maximum linear velocity of a cell would be $393\ \mu m/200\ ms = 2\ mm/s$ at 10X magnification and $197\ \mu m/200\ ms = 1\ mm/s$ at 20X. For slower velocities, we could increase the delay time for longer integration times and increased signal. We also note that beyond this proof-of-concept demonstration, an alternative CCD with faster readout could be implemented to increase maximum velocities at the cost of shortened integration times.

4.3.3 Optimization of TDI Synchronization and Bead Signal-to-Noise Ratio (SNR)

We optimized the TDI SFC system using calibration beads (6 μm) with relative fluorescence intensities from 0.3 to 100%. For a frame of reference, we compared these beads, which have an arbitrary intensity ranging from 0.3% to 100%, to antibody binding capacity (ABC) beads to find out the antibody loads for each bead population. ABC beads contain Fc-specific receptors and carry a known, quantitative load of fluorescent antibodies. Both calibration beads and ABC beads were measured by traditional flow cytometry on the same run (**Figure 4.4**).

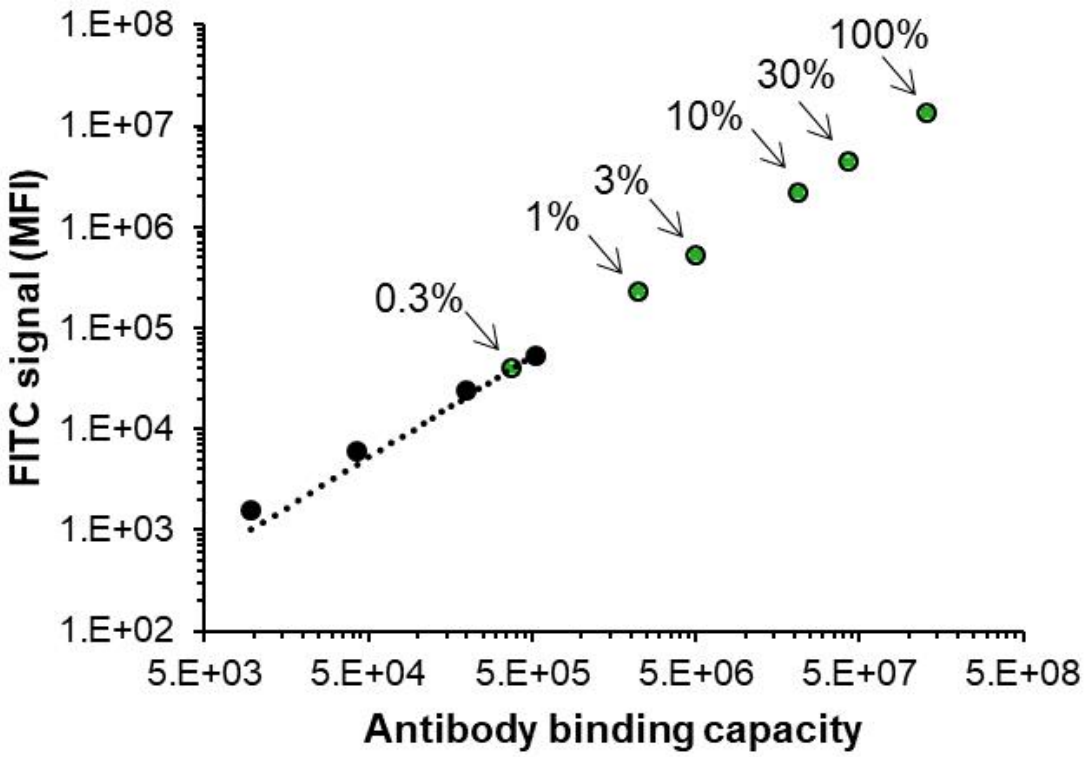


Figure 4.4 Labeled anti-mouse IgG Fc-specific beads with varying antibody binding capacity with mouse, anti-human TdT-FITC antibodies. Each population of ABC beads was measured using a traditional flow cytometer and the mean fluorescence intensity (MFI) was extracted. Fluorescence from the fluorescence calibration beads was then measured, and the calibration beads MFI values were correlated with the ABC values. The linear fit between MFI and ABC values is shown in black ($R^2 = 0.995$). A linear fit was also obtained between MFI and percent fluorescence for the calibration beads (data not shown, N=2, $R^2 = 0.996$). Note that we subtracted all positive populations with the MFI of their respective blank, unlabeled beads.

Only the 0.3% calibration beads were within the dynamic range of the ABC beads (3.8 $\pm 0.2\times10^5$ antibodies). However, assuming linearity beyond the ABC populations available, the 1% calibration beads would correspond to a load of 2.2 $\pm 0.1\times10^6$ antibodies. For a frame of reference, direct immunolabeled surface antigens typically fall in a range of 10^4-10^5 antibodies per biological cell. Therefore, the 0.3% to 100% fluorescence intensity corresponded to loads of 3.8 $\pm 0.2\times10^5$ to 1.3 $\pm 0.1\times10^8$ fluorescent antibodies, respectively (**Figure 4.4**). Because directly immunolabeled surface antigens typically fall in a range of 10^4-10^5 antibodies per biological cell, we aimed to achieve high sensitivity for the dimmest beads (0.3% intensity). Thus, we will

describe sequential improvements to the TDI SFC system's synchronization and fluorescence sensitivity (**Table 4.1**) before applying the optimized system to the immunophenotyping of a B-ALL cell line against the TdT leukemic antigen.

We note that a cell or bead will generate a TDI signal S(λ,t), which is a fluorescence spectrum dispersed by wavelength (λ) and read over time duration (Δt). With improved synchronization between cell velocity and TDI clocking, peak signal will be maximized, and event duration will be minimized (see peak shapes in **Figure 4.3**). Thus, for variables that affect TDI synchronization (**Table 4.1**), we monitored Δt and the peak signal S_{peak} observed over Δt at the maximum emission wavelength λ_{max}:

$$S_{peak} = \max\left(S(\lambda_{max}, t)\right) \qquad \text{(Eq. 4.1)}$$

For variables that only affect the system's fluorescence sensitivity, such as excitation power, we accumulated all integrated signal S_{int} throughout the event's duration and across all wavelengths:

$$S_{int} = \int_0^{\Delta t} \int S(\lambda,t)\, d\lambda\, dt \qquad \text{(Eq. 4.2)}$$

We also calculated the SNR for each event by dividing the accumulated signal by noise in the readout. To account for both dark noise, read noise, and shot noise, we fit the integrated spectrum with a smoothing spline function $F_{fit}(\lambda)$ (**Figure 4.5**) and calculated noise from the square root of the residual's sum of squared errors (SSE):

$$SNR = \frac{S_{int}}{\sqrt{SSE}} = S_{int} / \sqrt{\sum \left(\int_0^{\Delta t} S(\lambda,t) dt - F_{fit}(\lambda) \right)^2} \qquad \text{(Eq. 4.3)}$$

A TDI event consists of a fluorescence spectrum read over time, S(λ,t). To assess the effect of the variables in **Table 4.1** on fluorescence sensitivity, we accumulated all signal collected by the CCD by integrating the total signal S_{int} over all wavelengths and over the event's

duration Δt. We then evaluated noise in each fluorescence spectrum integrated over time, i.e., $\int_0^{\Delta t} S(\lambda,t)dt$, by fitting a spline function $F_{fit}(\lambda)$.

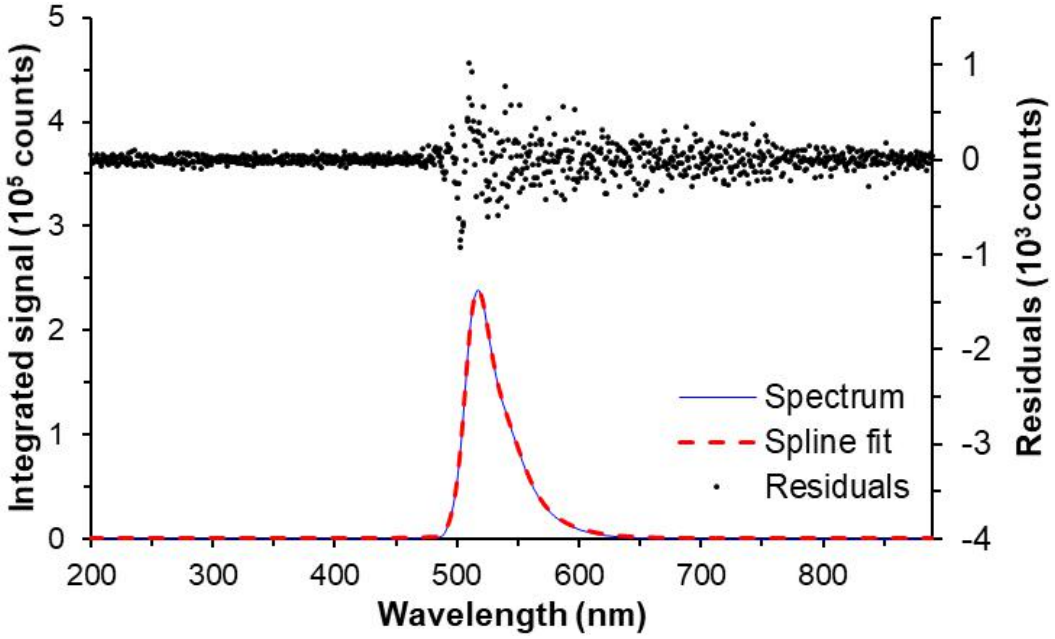


Figure 4.5 An integrated spectrum was fit with a spline function (smoothing parameter = 0.1). Residuals in the fit represented noise in the integrated signal, quantified by the sum of square errors (SSE).

The residuals from the fitting function represent noise in the integrated spectrum that was quantitated by the sum of square errors (SSE) over all wavelengths. We note that we did not observe a dependence on the signal-to-noise ratio on the wavelength region used to sum the noise. An example of the fitting process is shown in **Figure 4.5**, where an increase in shot noise is evident as the fluorescence signal increases.

Table 4.1 Variables tested for the TDI SFC system that improved TDI synchronization and/or fluorescence sensitivity. The related data are summarized by median improvement to S_{peak} or SNR.

TDI SFC variable	Range	Improvement	Figure
Cell focusing	None → 1D focusing	Synchronization: 4.5× S_{peak}, 0.8× Δt	4.6
Cell velocity	1.41 → 1.83 ← 2.11 mm/s	Synchronization: 2.3× S_{peak}, 0.5× Δt	4.7
Excitation power	4 → 21 mW	Sensitivity: >9× S_{int}	4.8
Objective NA	0.5 → 0.75 NA	Sensitivity: 2× S_{int}, 1× SNR	4.9, 4.10
TDI integration time	200 → 400 ms	Sensitivity: 8-10× S_{int}, 4× SNR	4.11

4.3.3.1 One Dimensional Microfluidic Focusing for Improved TDI Synchronization

Hydrodynamic flow generates non-uniform velocities throughout a capillary or microchannel's cross-section, which can cause variable TDI synchronization from cell-to-cell. We observed this effect by infusing fluorescent beads through a 75 μm square, glass capillary at an average velocity of 2 mm/s with a 200 ms integration time. Note that in all cases, we verified the TDI SFC focal plane was focused to the center of the capillary or microchannel using the focusing system and a micrometer sample mount. We counted TDI signals when the collected spectrum increased 20× above background noise and analyzed events for time duration Δt and S_{peak} at the bead's emission maximum (515 nm). In unfocused hydrodynamic flow, TDI signals ranged from 24-124 ms in duration, and many events lacked a peak signal characteristic of synchronized TDI and instead exhibited flat, constant signal (**Figure 4.6B**).

To improve velocity uniformity, we employed a microfluidic flow cell with a T-junction that merged sheathing buffer on either side of the sample stream to provide 1D focusing along the lateral, x-direction of the flow cell (**Figure 4.6A**). This compressed the sample stream into a more uniform lateral position, thereby improving velocity uniformity through the microchannel's cross-section (**SI Figure 4.4A**). Without optimizing cell velocity to CCD clocking (average velocity 1.6 mm/s, maximum 2.4 mm/s), we observed more uniform event durations, most frequently at 64 ms (**Figure 4.6C**) with peak shapes characteristic of TDI synchronization. Indeed, S_{peak} increased by 4.6× with microfluidic 1D focusing (**Figure 4.6D**). Also, the event lengths distribution of 1D focusing showed higher uniformity than no-focusing (**Figure 4.6C**).

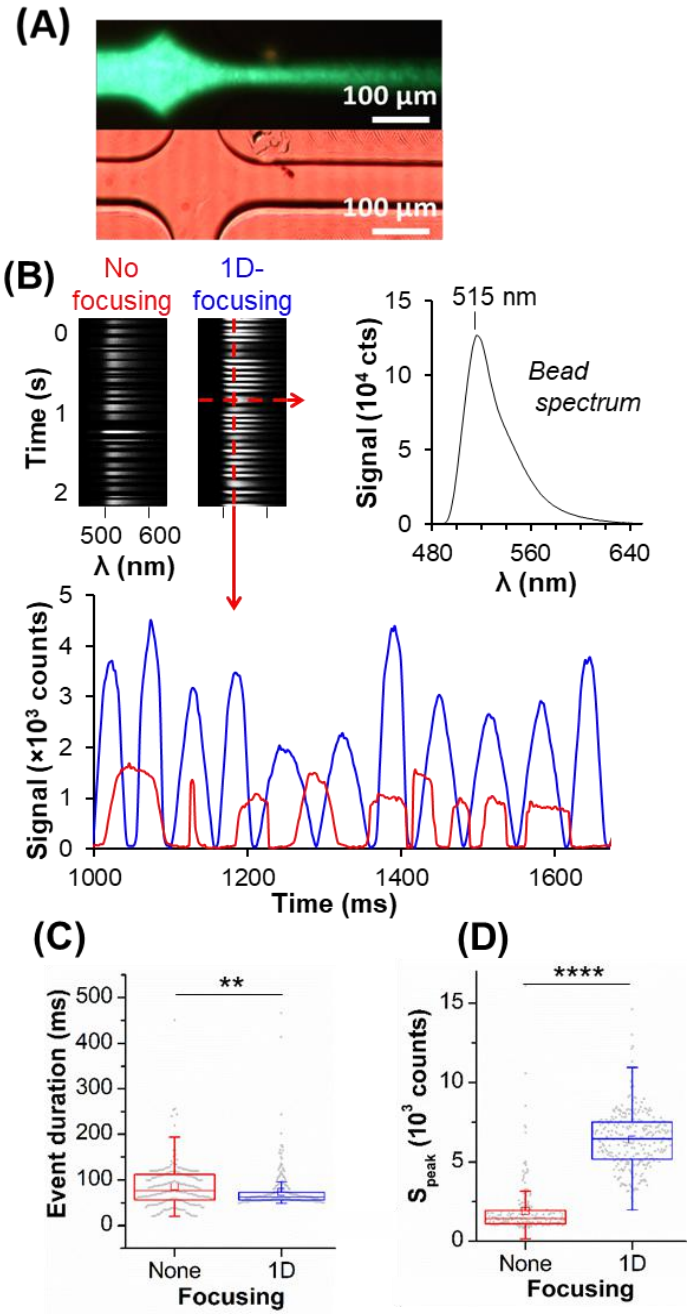


Figure 4.6 (A) A bright field image of the T-shaped microfluidic flow cell and a fluorescent image showing hydrodynamic focusing in the flow cell using 50 mM FITC solution and deionized water. (B) TDI imaging of fluorescent beads (100%) without focusing (red) or with 1D microfluidic focusing (blue). Shown are spectra acquired over 2 s with background data between events removed. We plotted a bead spectrum along the CCD's wavelength axis and the signal at λ_{max} (515 nm) over time. Box plots of **(**C**)** event duration and **(**D**)** S_{peak}. In box plots: middle line shows median; edges of box, upper and lower quartiles; bars, minima and maxima excluding outliers; squares, averages; and dots, all data. Statistical analyses showed $p<10^{-2}$ (**) and $p<10^{-4}$ (****).Sample and sheath flow rates were 0.178 and 0.217 uL/min, respectively, and 10X objective with 4 mW excitation was used for both data sets.

We then optimized flow conditions for 1D microfluidic focusing by maintaining the sample flow rate and increasing the sheathing rate. Thereby, we increased maximum velocity in the sheathed sample stream from 1.41 to 2.11 mm/s while decreasing the width of the sheathed sample stream from 24 μm to 16.4 μm (**SI Figure 4.4B**). Examples of a well-synchronized and a bad-synchronized bead event are shown in **Figure 4.7A**. We maintained a constant integration time of 200 ms and observed an optimal velocity (1.83 mm/s) that maximized peak signal (median, 13,430 counts) and minimized event duration (median, 36 ms) with statistical confidence that the observed differences were not due to sampling error ($p<10^{-4}$, **Figure 4.7B**). These trends agree within 10% of theoretical calculations and illustrate that as TDI synchronization is optimized, signals are read as narrower, more intense peaks (**Figure 4.3)**.

Lastly, we note that the TDI SFC optical system does not use a confocal pinhole to block out-of-focus fluorescence. Cells not focused along the microchannel's 93 μm depth can be detected but will have lower signal, variable velocity, and imperfect synchronization (**SI Figure 4.5**). In future renditions of the TDI SFC system, we will employ 2D focusing to improve upon the gains made by 1D microfluidic focusing herein.

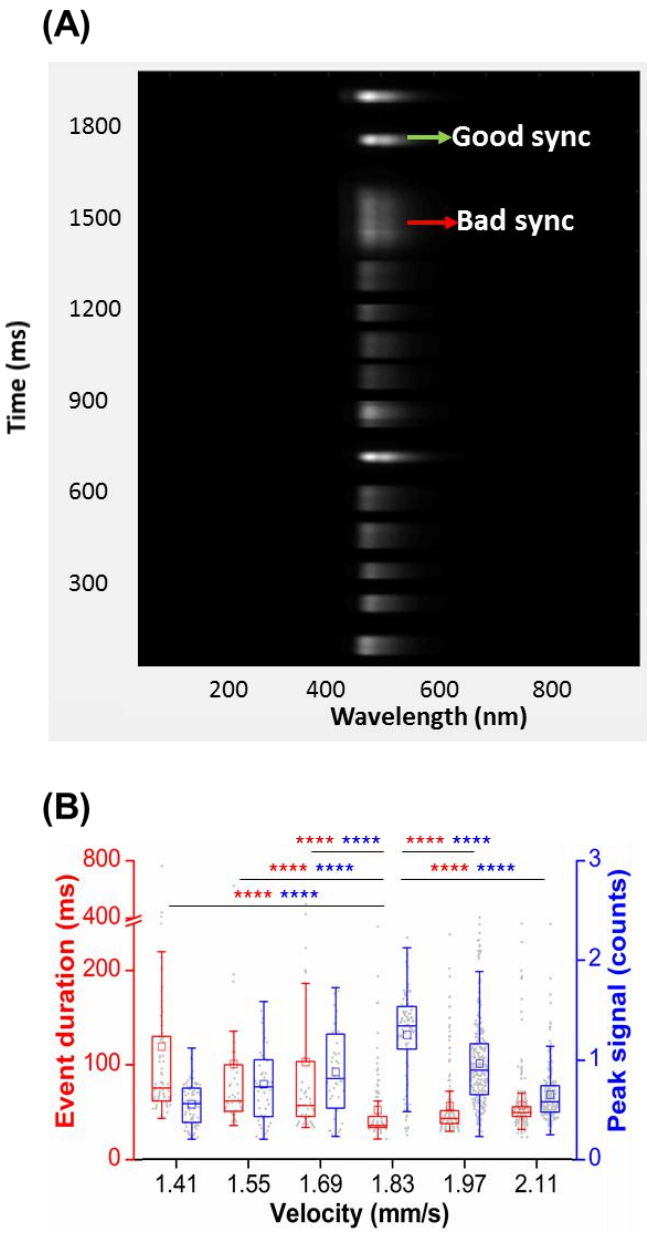


Figure 4.7 (A) Examples of a well-synchronized and a bad-synchronized events. The well-synchronized event showed relatively short time duration and bright fluorescence while the bad-synchronized event was a "slur" and showed dim fluorescence. (B) Flow optimization of TDI synchronization in the 1D microfluidic flow cell, showing box plots of event duration (red) and peak signal at 515 nm (blue) at varying maximum velocities. Data acquired with 10X objective and 4 mW excitation. Statistical analyses showed $p<10^{-4}$ (****).

4.3.3.2 Optimization of TDI SFC Sensitivity for Weak Fluorescence Signals

All experiments to optimize TDI flow synchronization used beads with 100% relative fluorescence intensity. These beads correspond to a load of 10^8 fluorescent antibodies, 1,000-

10,000× greater than expected for biological cells. Therefore, we evaluated other methods to increase the TDI SFC's sensitivity.

4.3.3.2.1 Optimization of Excitation Power

We first increased the excitation power in the objective's field-of-view from 4 to 21 mW (**Figure 4.8A**). We enlarged a 488 nm excitation laser (54 mW) with a beam expander then used a circular aperture to cut the edges of the Gaussian laser beam. The circular aperture was closed (diameter d = 4.7 mm) to eliminate illumination outside the CCD's sensors field-of-view, thereby establishing Köhler epi-illumination. This also served to reduce non-uniformity in excitation power density through the field-of-view by eliminating light on the edge of the Gaussian beam's profile.

After beam expansion, the beam's radius (w = 1/e2 = 0.65 mm) was enlarged by a factor of 7.1× and 21.2× to w = 4.6 mm or 13.8 mm, respectively. With both beam expanders, the total beam power P_0 = 54 mW remained constant, but as the beam expanded, the laser power density became more uniform while decreasing in magnitude (**Figure 4.8A**). The profile of the enlarged beams is given by a Gaussian function over radius r:

$$I(r) = \frac{2P_0}{\pi w^2} e^{-\frac{2r^2}{w^2}} \qquad \text{(Eq. 4.4)}$$

Integration of Eq. 4.4 over the aperture's fixed surface area, $\pi d^2/4$, shows that the total power should increase from 5.7 mW to 40.6 mW when decreasing the beam size with 21.2× and 7.1× beam expanders. These theoretical calculations agree well with the experimental measurements of 4 mW and 21 mW, respectively. Lastly, we note that while the power density non-uniformity increases to 15% (RSD) with the 7.1× beam expander, even the lowest intensity excitation at the edge of the field-of-view surpasses the power density achieved with 21.2× beam expander (**Figure 4.8A**).

The signal at λ_{max} from the 100% intensity beads increased beyond the CCD's dynamic range (**Figure 4.8B**). Thus, we evaluated S_{int} at 550 nm, where the CCD was not saturated, and observed a 9.8× increase in S_{int} (5.2×10^6 to 4.7×10^7 counts for 4 to 21 mW, respectively).

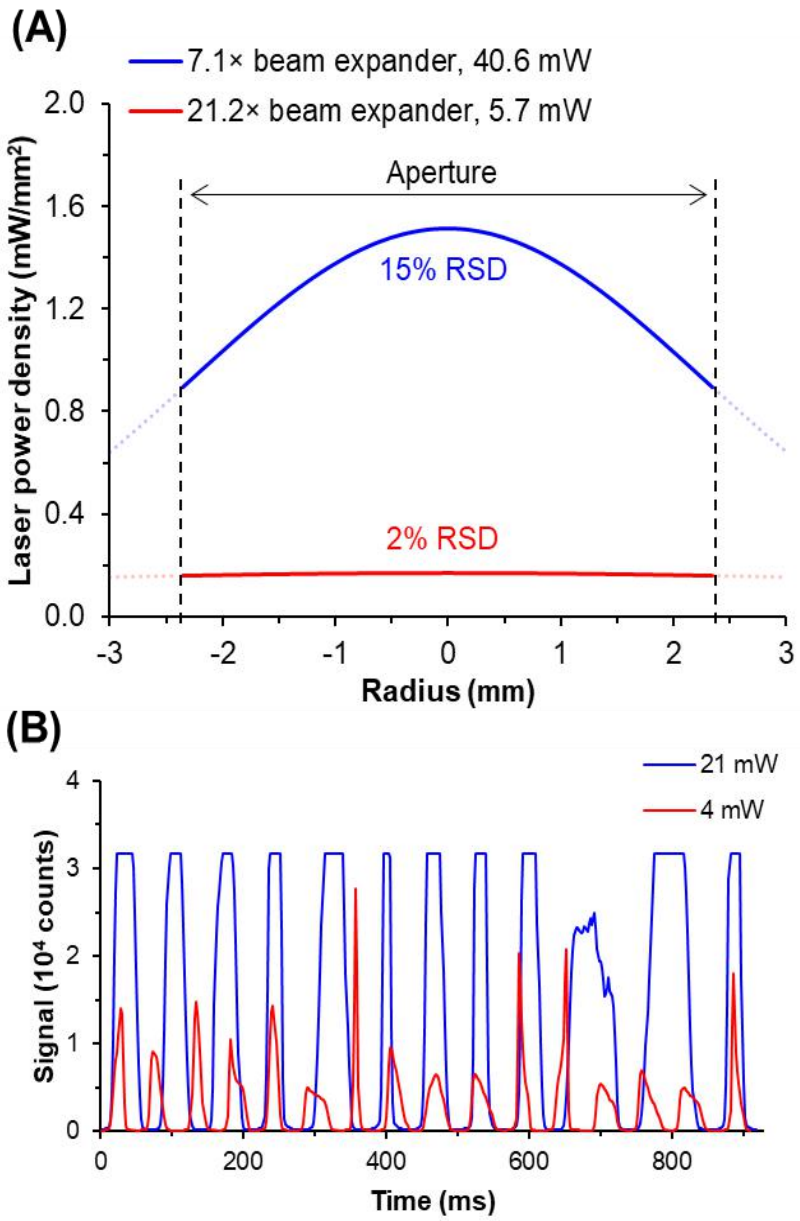


Figure 4.8 (A) Theoretical power density of the 488 nm excitation laser through the epi-illumination aperture for the 7.1× and 21.2× beam expanders. (B) Signals for 100% intensity beads at 4-21 mW excitation.

4.3.3.2.2 Higher NA objective

Given the enhanced signal with increased excitation power, we evaluated dimmer beads with relative intensities of 0.3% to 10% (corresponding to 3.8×10^5 to 5.0×10^6 Abs, **Figure 4.4**) with the 10X objective (0.50 NA). We observed a linear increase ($R^2 = 0.99$) in median S_{int} from 1.0×10^5 to 8.3×10^6 counts from 0.3% to 10% intensity (**Figure 4.9A**) and a 152 SNR (median) for the dimmest, 0.3% intensity beads (**Figure 4.9B**).

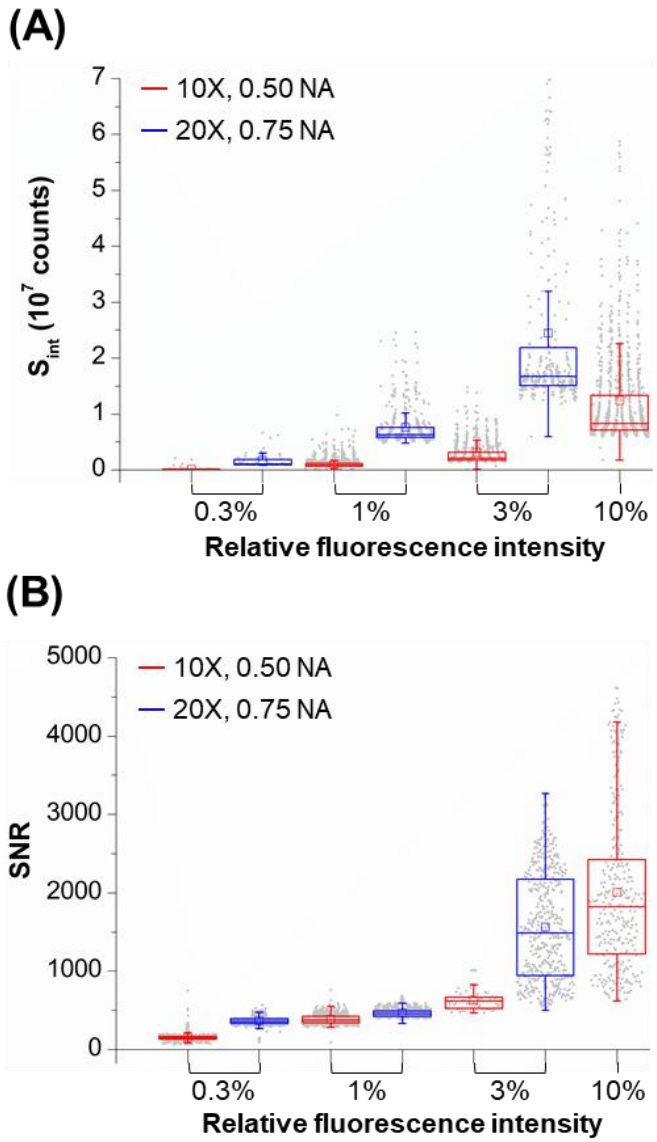


Figure 4.9 (A**)** S_{int} and **(**B**)** SNR for 0.3%-10% beads at 10X (0.50 NA) or 20X (0.75 NA), 21 mW.

We repeated these experiments with a 20X objective with a higher NA (0.75) to increase the amount of emission gathered. We observed a substantial increase in signal and could not analyze the 10% intensity beads due to CCD saturation. For 0.3% to 3% intensity beads, the median S_{peak} increased by 7.5-10.2× (**Figure 4.9A**) and SNR values increased by 4.0-4.8× (**Figure 4.9B**), respectively. The dimmest 0.3% beads produced a median S_{int} of 1.1×10^6 counts (**Figure 4.9A**) at a SNR of 610 × (**Figure 4.9B**), and median S_{int} again increased linearly with bead intensity (R^2=0.99) to a median SNR of 1,820 for 3% intensity beads.

As an example, the substantial signal increase on 1% relative intensity beads is shown in **Figure 4.10**. As a note, we included only data that surpassed a threshold of 20× the signal-to-background noise. We also removed all background data between events, so we could show several events on the same plot.

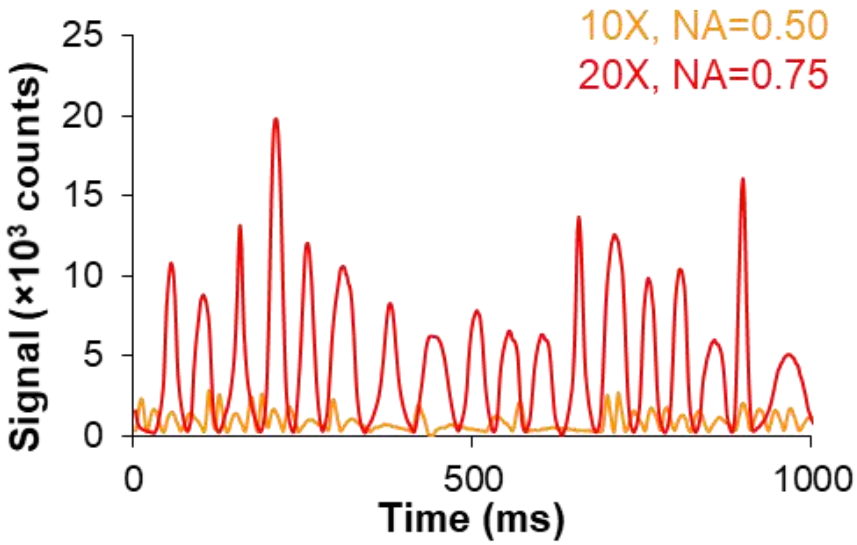


Figure 4.10 TDI signals at 515 nm for the 1% intensity calibration beads, integrated for 200 ms using a 10X objective (NA=0.50, 1.83 mm/s) or a 20X objective (NA=0.75, 0.92 mm/s). Statistical analysis is provided in **Figure 4.9**.

4.3.2.2.3 Higher full-frame integration time

The above data represent substantial gains in sensitivity for fluorescent calibration beads that approach the expected load of Abs on immunolabeled cells, but these gains come at a cost of

reduced sample processing rate (i.e., throughput). For example, increasing the NA of the collection optic reduced the field-of-view by a factor of 2. Given a constant integration time (200 ms), we had to reduce all flow rates by ½ to maintain an optimized velocity of 0.92 mm/s for TDI synchronization (**Figure 4.10**).

In a similar manner, we halved the sample velocity when doubling the integration time from 200 ms to 400 ms (**Figure 4.11**). The 1.97× increase in S_{int} resulted in only a minimal change in SNR ($p = 0.071$). Thus, when analyzing cells, we chose to use the 20X objective (0.75 NA), 200 ms integration time, and 21 mW excitation power.

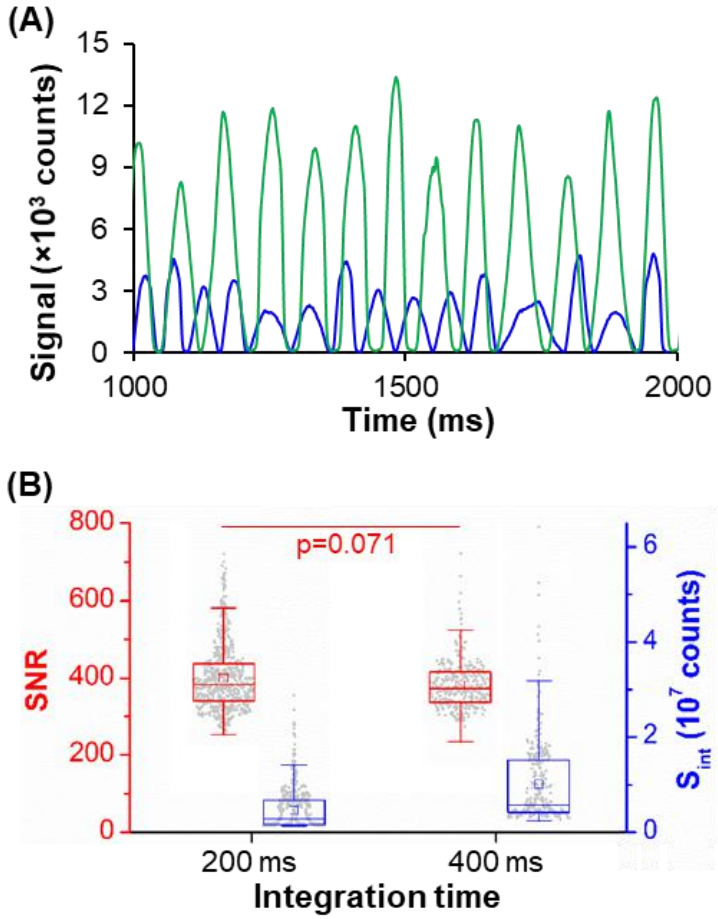


Figure 4.11 (A) TDI signals at 515 nm for the fluorescent calibration beads integrated for 200 ms (2.0 mm/s) or 400 ms (0.92 mm/s) using 100% intensity calibration beads and a 10X objective. **(**B**)** Box plots for increased integration time, 3% beads at 10X. Unless noted, statistical analysis showed $p<10^{-6}$.

4.3.4 TDI SFC Detection Efficiency and Throughput

To evaluate detection efficiency, we processed concentrations of fluorescent beads (3% intensity, 10X objective, 200 ms integration) from 200-1,500 beads/µL and measured the bead concentration by dividing the number of beads detected by the sample volume processed. We note that there is no limit for lower bead concentrations; we designed these experiments to test the upper limit of throughput. We correlated results to a commercial MFC (BD Accuri C6 Plus). From 200 to 750 beads/µL, we observed high concordance ($R^2 = 0.97$) between the systems in terms of the number of beads detected (**Figure 4.12** ***left***).

At concentrations above 750 beads/µL, TDI SFC detection efficiency decreased, and the number of detected events remained fairly constant. Upon inspection of TDI signals, a substantial number of events contained multiple, unresolved beads overlapping in time (starred in **Figure 4.12** ***right***).

BD Accuri C6 Plus (beads/µL)	TDI-SFC (beads/µL)	TDI-SFC efficiency
205 ±5	203 ±10	95 ±5%
303 ±13	226 ±34	75 ±12%
665 ±55	698 ±60	105 ±13%
745 ±11	692 ±8	93 ±2%
746 ±39	763 ±40	102 ±8%
1,353 ±29	770 ±74	Out of TDI-SFC range
1,449 ±103	931 ±95	

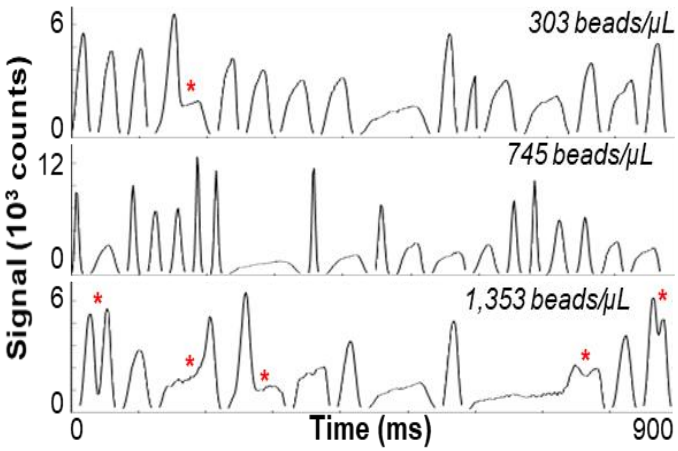
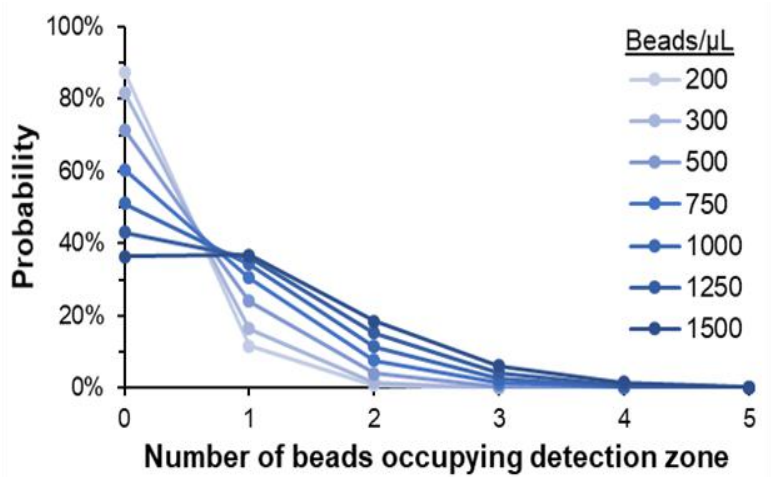


Figure 4.12 TDI SFC measurements of bead concentration (3% intensity, 10X objective, 200 ms integration) compared to a commercial MFC instrument. (*left*) TDI SFC detection efficiency compared to commercial MFC. (*right)* At high bead concentration, coincident beads were not resolved (noted as red stars).

To calculate the probability of occupancy in the TDI SFC field-of-view, we calculated the detection volume (673 pL) using the sheathed sample width (18.6 µm, **SI Figure 4.4B**) and assumed that beads can be detected throughout the microchannel's depth (92 µm) and the CCD's

field-of-view (393 μm). We calculated the probability that k beads would occupy the field-of-view simultaneously (**Figure 4.13**) by a Poisson distribution (P(k)=e^(-λ) λ^k/k!, where λ is the average occupancy (bead concentration multiplied by the detection volume). The probability of coincident events based on Poisson statistics (**Figure 4.13**) were calculated to be 1%, 10%, and 27% for 200, 750, and 1,500 beads/μL, respectively. This indicates a upper throughput of 90 beads/min (750 beads/μL) in the current TDI SFC system.



Beads/μL	Probability of multiple occupancy
205	0.8%
303	1.8%
665	4.5%
745	9.2%
746	14.6%
1353	20.6%
1449	26.7%

Figure 4.13 Poisson distributions showing multiple bead occupancy over bead concentrations relevant to **Figure 4.12**.

TDI operation should be capable of resolving coincident events, but the current 1D microfluidic focusing strategy does not focus events along the z-direction. Beads with varying velocities and slurring from imperfect synchronization can lead to event overlap. Variability in the z-axis travel of beads also affects the consistency in an event's fluorescence intensity. This can increase the coefficient of variation (CV) for a single bead population.

We measured 3% intensity beads with a commercial MFC and by TDI SFC and measured signal as the mean fluorescence intensity (MFI) or S_{int}, respectively (**Figure 4.14**). Commercial MFC yielded a narrow distribution centered at 5.4×10^5 counts with a 5.7% coefficient of variation (CV = MFI/standard deviation). Integrated TDI signals were non-Gaussian, skewed

because slower beads are imaged by the CCD longer and generate increased S_{int} even though S_{peak} decreases. While the non-Gaussian distribution makes the traditional CV inappropriate, we report an average S_{int} of 2.2×10^6 counts with a traditional CV of 69.1%. More appropriately, we calculated the CV based on the median signal divided by the interquartile range (refer to box plot inlaid in **Figure 4.14**). This yielded a median S_{int} of 1.9×10^6 and a reduced CV of 24.2%. The relatively high variation in signal by TDI SFC can be attributed to variability in bead position in the z-axis focal plane using 1D microfluidic focusing.

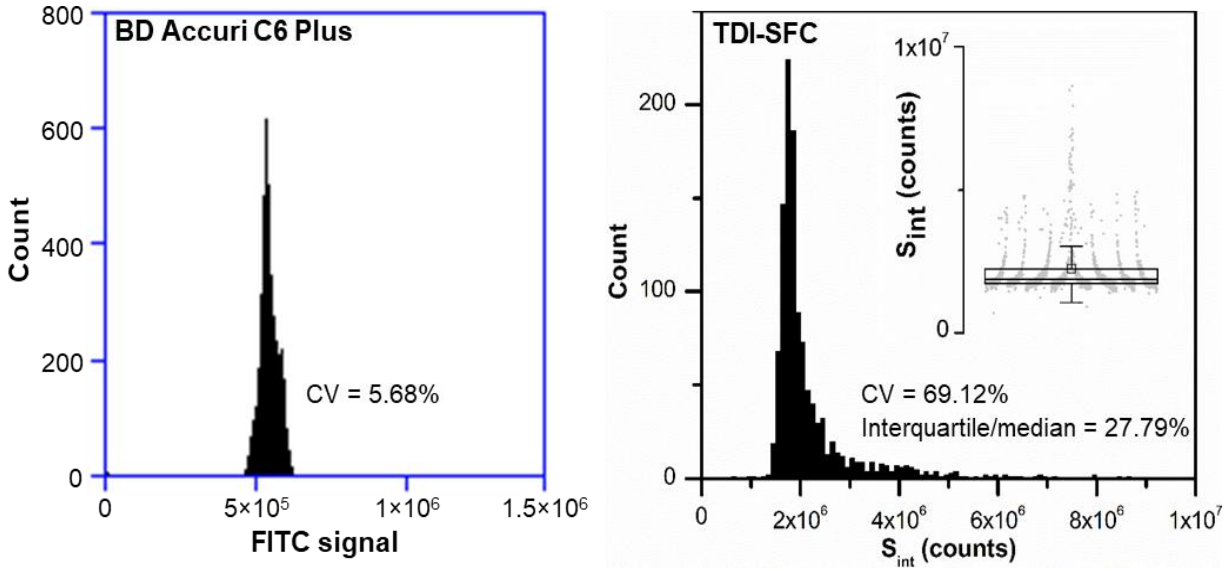


Figure 4.14 Histograms of fluorescence intensity for 3% intensity beads measured with a commercial MFC and TDI-SFC (10X objective, 200 ms integration). For the TDI-SFC data, a box plot is inlaid to show skew in the signal distribution that led to using the interquartile distance divided by median value for calculating the CV.

4.4 Conclusion

A unique TDFI-SFC instrument was reported for the automated immunophenotyping of rare, low abundant biological cells. CCD readout of high-resolution fluorescence spectra was conducted in a TDI mode, ensuring that rare cells would not be missed because the sensor duty cycle approached 100% and the fluorescence intensity was high even for weak emitting events. The TDI SFC could correctly identify rare cells due to the ability to acquire the entire

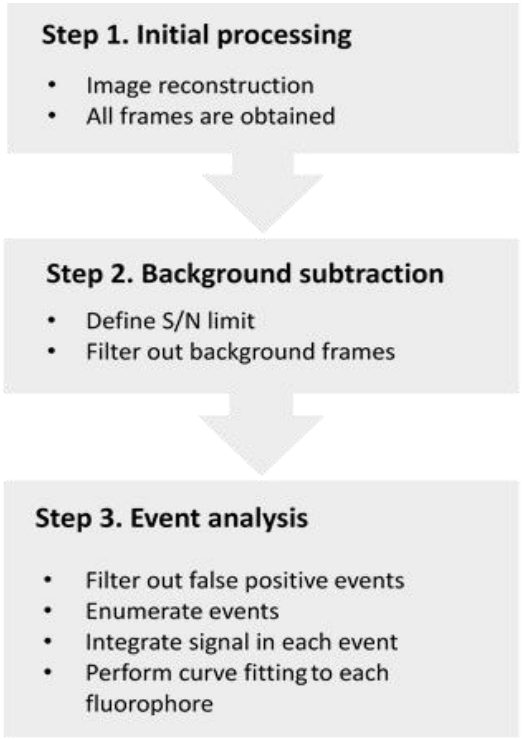
immunolabels' fluorescence emission spectrum to allow for spectral deconvolution as opposed to conventional MFC, which typically requires >20,000 events to set gating thresholds for cluster analysis. This attribute will be particularly important for enumerating CLCs for monitoring MRD because the biological specificity of the affinity agent is modest and the number of cells enriched makes it operationally difficult to use fluorescence microscopy.

In this chapter, the TDI SFC was characterized using a biological cell model, fluorescent beads. A microfluidic flow cell with T-geometry of channels was fabricated and used as a hydrodynamic flow focusing device. To validate the benefit of 1D hydrodynamic focusing in the microfluidic flow cell, data comparison to non-flow focusing was undertaken. As expected, using 1D hydrodynamic focusing produced higher fluorescence signal and tighter event length distribution. We also used the fluorescent beads and tried multiple flow rates in the T-shaped microfluidic chip to find out the optimal flow rate for TDI synchronization. And it suggested that a linear velocity of 1.83 mm/s of the beads provided the best synchronization. To further improve the sensitivity, optimizations such as using higher power excitation source, using a higher NA objective and adding a delay time to the TDI CCD were performed. These optimizations successfully increased the fluorescent signal by 9× S_{int}, 2× S_{int} with 1× SNR and 8-10× S_{int} with 4× SNR, respectively. The dynamic range of throughput was also explored by testing the same bead population with various concentrations in our TDI SFC and a commercial MFC. The results indicated a upper throughput of 90 beads/min (750 beads/μL) in the current TDI SFC system. All of these characterizations were necessary and critical to facilitate successful experiments on analyzing real biological samples, such as leukemic cells or epithelial cancer cells.

4.5 Appendix-Supplemental Information

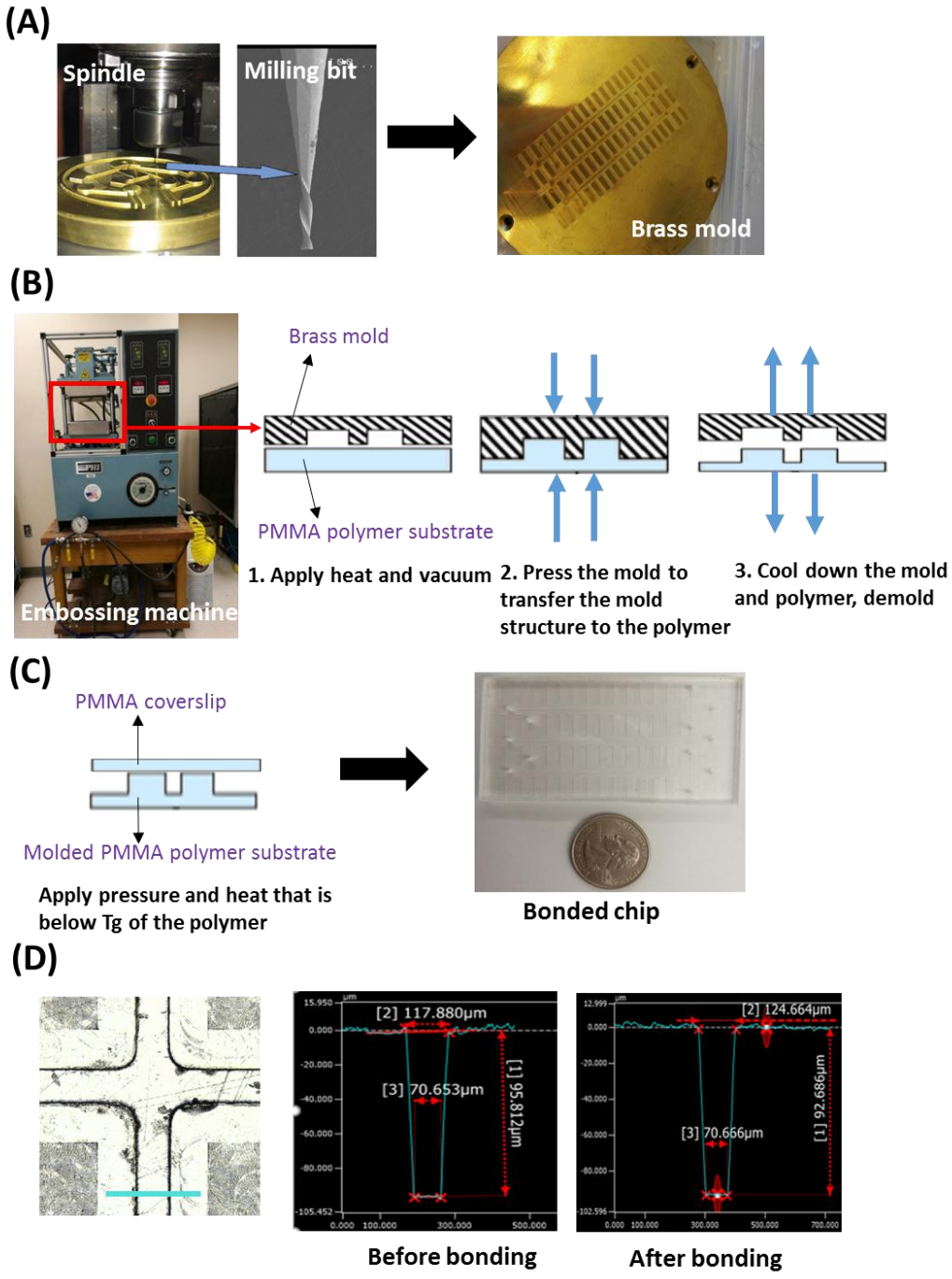
Table of Contents

1. MATLAB data processing workflow



SI Figure 4.1 MATLAB data processing steps and the detailed processes in each step.

2. Fabrication, thermal fusion bonding and channel dimension characterization of the microfluidic flow cell.



SI Figure 4.2 Fabrication and channel characterization of the microfluidic flow cell. (A) Produce the brass mold by high-precision micromilling. (B) Transfer the mold structure to the PMMA polymer substrate by hot embossing. (C) Use thermal fusion bonding to plug the microfluidic flow cell. (D) Use a laser-scanning confocal profilometer to characterize the microfluidic channel dimensions before and after thermal fusion bonding to ensure proper structure fidelity.

3. Measurement of the minimum delay time of our TDI SFC.

The delay time per pixel row can be expressed as:

$$t_{delay} = N_x(t_{st} + t_v) + t_i + t_d \quad \text{(Eq SI 4.1)}$$

Thus, the full frame integration time t_{int} of TDI can be expressed as:

$$t_{int} = t_{delay} * N_y = [N_x(t_{st} + t_v) + t_i + t_d]N_y \quad \text{(Eq SI 4.2)}$$

Where N_x is the number of pixels at spectral axis (N_x =1340 in this study), t_{st} is the time it takes to shift one pixel out of the serial register, t_v is the time it takes to digitize one pixel, t_i is the time it takes to shift the photoelectrical signals from one pixel row to the next pixel row, t_d is the secondary delay time that can be manually set on the Labview software user interface of our TDI SFC, and N_y is the number of pixels on shift direction (N_y=100 in this study).

The minimum t_{delay} is when t_d is set as 0. Despite of the expression of the minimum delay time per pixel row, it is unnecessary to measure t_{st}, t_v or t_i respectively to know t_{delay}. Instead, t_{delay} can be calculated by collecting TDI data. Specifically, if the data acquisition time is recorded and the number of frames is obtained from data processing in MATLAB, the minimum t_{delay} per pixel row can be calculated as:

$$t_{delay} = \frac{t_a}{N_a} - t_d \quad \text{(Eq SI 4.3)}$$

Where t_a is the TDI data acquisition time, N_a is the number of frames acquired and it can be obtained after processing the data in MATLAB, and t_d is the secondary delay time that can be manually set on the Labview software user interface of our TDI SFC. The measurements of the minimum t_{delay} per pixel row are shown in **SI Table 4.1**.

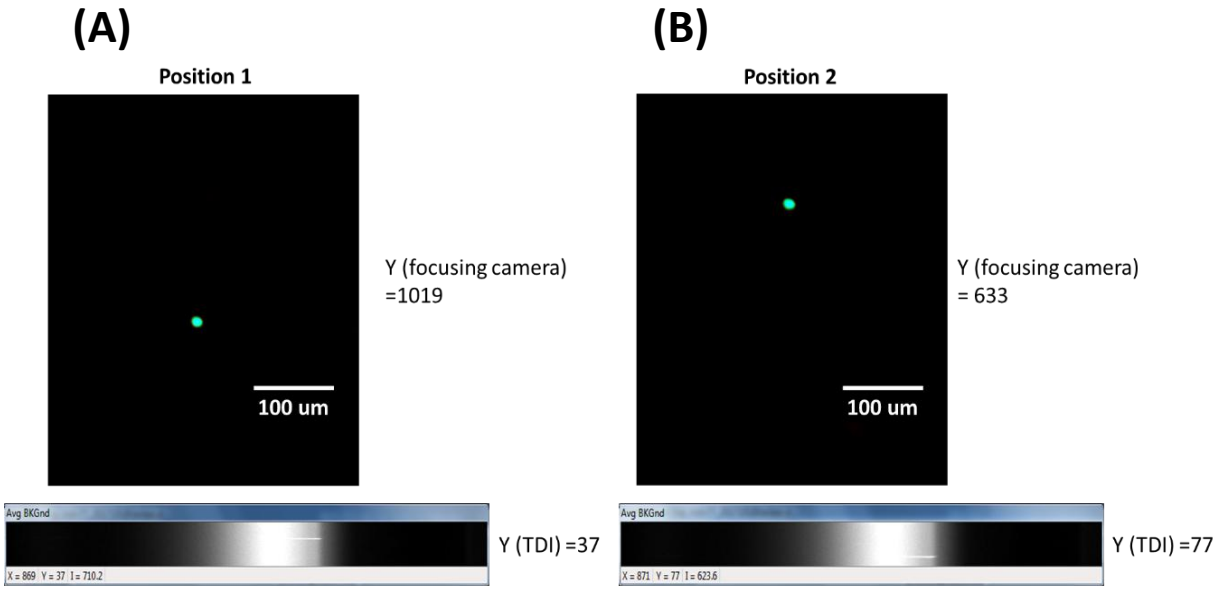
SI Table 4.1 Measurements of the minimum t_{delay} per pixel row.

Trail number	Data acquisition mode	Binning	Secondary delay time (s)	Data acquisition time	Data acquisition time (s)	Total number of frames acquired	Minimum delay time per pixel row (s)	
1	TDI	NA	0.0021	5m9s	309	77280	0.00190	
2	TDI	NA	0.0022	5m13s	313	75420	0.00195	
3	TDI	NA	0.0023	5m2s	302	71863	0.00190	
4	TDI	NA	0.0024	4m35s	275	63595	0.00192	
5	TDI	NA	0.0025	5m11s	311	70587	0.00191	
6	TDI	NA	0	6m5s	365	186682	0.00196	
7	TDI	NA	0	5m8s	308	156058	0.00197	
8	TDI	NA	0	5m52s	352	170460	0.00207	
9	TDI	NA	0	5m23s	323	161980	0.00199	
10	TDI	NA	0	6m17s	377	185854	0.00203	
							0.00196	**Average**
							0.00006	**STDEV**

From our measurements, the minimum t_{delay} per pixel row was 0.002s. Thus, the minimum full frame integration time t_{int} for the 100 pixel rows was 0.2 s.

4. Measurement of FOV with a 10x objective

The FOV was measured by mounting a glass slide containing fluorescent calibration beads sealed with a glass coverslip. It was measured with reference of the CMOS alignment camera. An example measurement is shown in **SI Figure 4.3**. By recording the full frame snapshots from the CMOS alignment camera and our CCD camera of one bead at two different vertical positions (Y axis in **SI Figure 4.3**), the FOV could be measured.



SI Figure 4.3 (A) When the bead was in position 1 it corresponded to a position of Y=1019 pixels on the focusing camera and a position of Y=37 pixels on the full frame image of TDI CCD. (B) After changing the bead's position to position 2, it corresponded to a position of Y=633 pixels on the focusing camera and a position of Y=77 pixels on the full frame image of TDI CCD.

The bead's position difference on the focusing camera was 1019-663=356 (pixels). Through a calibration slide on the focusing camera we could get the pixel calibration information of 0.4338 μm/pixel. Thus, the physical position difference between position 1 and position 2 could be calculated as:

$$356 \text{ pixels} \times 0.4338 \text{ μm/pixel} = 154.4 \text{ μm}.$$

The TDI CCD had 100 pixels on Y axis (time axis or shift direction). Since the bead's physical position difference resulted in a difference of 77-37=40 (pixels) on the full frame images on TDI, the FOV on the time axis could be calculated as:

$$\text{FOV} = \frac{154.4 \text{ μm}}{40 \text{ pixels}} \times 100 \text{ pixels} = 390 \text{ μm}$$

Three trials of the measurements are shown in **SI Table 4.2**:

SI Table 4.2 Measurements of FOV

Trial 1	Trial 2	Trial 3	AVG
390 μm	389 μm	401 μm	393.3μm

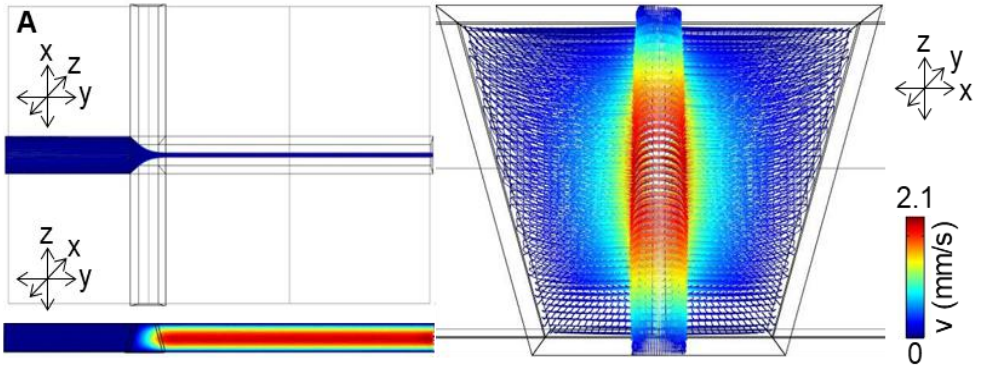
5. COMSOL fluid dynamic simulations for the microfluidic flow cell

We evaluated the flow velocity profiles within the microfluidic T-cell with hydrodynamic focusing using 3-D laminar flow simulations in COMSOL Multiphysics 5.2a. We constructed the geometry using measurements of the microchannel cross section obtained from laser scanning profilometry, which we provide along with all other parameters of the simulations in **SI Table 4.3**. We evaluated the flow profile in the square bore capillary with no hydrodynamic focusing using the same simulation parameters, except that the geometry was designed as a 1 mm long segment of a 75 μm square channel. We show axial cross sections of the velocity flow profiles in the sheathed sample stream in **SI Figure 4.4A** with the maximum velocities and core widths noted. Results for the square capillary are not shown, but these results indicated a maximum velocity of 4.19 mm/s and average velocity of 2 mm/s for the volume flow rate tested, which was 0.675 μL/min. We also note that we observed variation in the maximum wavelength of bead spectra during TDI SFC imaging, which correlates with the core width of the sheathed sample stream (**SI Figure 4.4C**), because the spatial x-axis is spectrally sorted into the wavelength axis of the TDI SFC CCD (see **Figure 4.2**).

SI Table 4.3 All simulation parameters for 3-D COMSOL Multiphysics simulations of velocity fields.

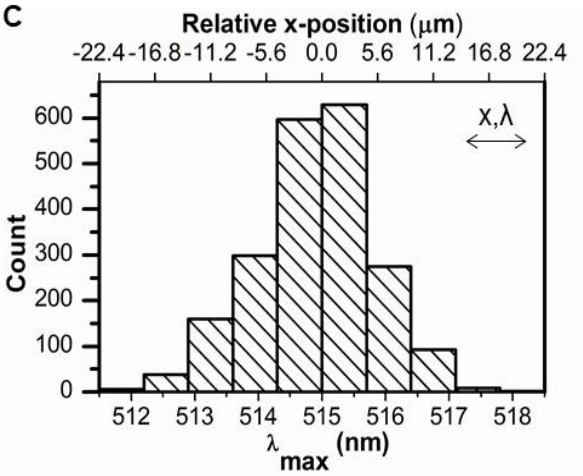
Simulation parameter	Value
Geometry lengths	1.5 mm (sample), 1 mm (both sheaths)
Geometry height	92.6 μm
Geometry width (top)	124.6 μm
Geometry width (bottom)	70.6 μm
Fluid properties	Water
Wall conditions	No slip
Inlet condition	0.12 μL/min (sample), 0.12-0.21 μL/min (sheath)
Outlet condition	0 Pa
Initial values	0 velocity field, 0 pressure drop
Meshing preset	Physics-controlled, “Extra fine”
Number of elements	2,823,023
Element size	0.456-7 μm

Maximum element growth rate	1.08
Resolution of curvature	0.3
Resolution of narrow regions	0.95
Number of boundary layers	2
Boundary layer stretching factor	1.2
Thickness adjustment factor	5
Solution type	Stationary
Solver	GMRES, fully coupled (default settings)
Relative tolerance for convergence	10^{-3}



B

Sheath rate (µL/min)	Core width (µm)	V_{max} (mm/s)	V_{ave} (mm/s)	RSD
0.120	24.0	1.41	0.91 ±0.41	
0.138	22.0	1.55	1.01 ±0.45	
0.156	20.2	1.69	1.10 ±0.50	45%
0.174	18.6	1.83	1.20 ±0.54	
0.192	17.6	1.97	1.29 ±0.58	
0.210	16.4	2.11	1.39 ±0.62	
None	92.0	0.47	0.24 ±0.16	66%

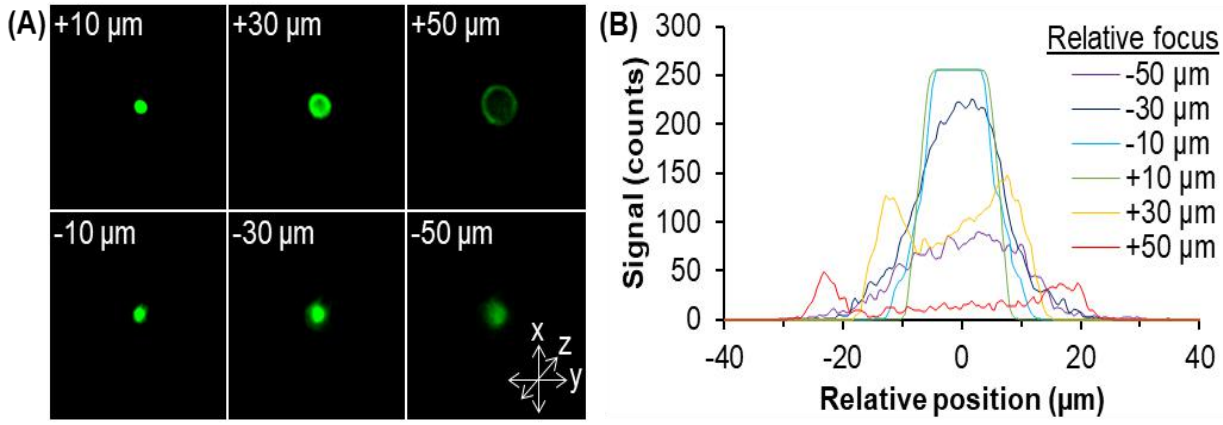


SI Figure 4.4 (A) 3-D COMSOL simulation results for the microfluidic T-shaped flow cell with hydrodynamic flow focusing showing streamline plots viewed from the top (top-left panel), side (bottom-left panel), and cross-section (right panel) for a sheath flow rate of 0.210 µL/min. (B) Table showing the maximum velocity and core width of the sheathed sample stream for the sheathing rates tested in the main text. No sheathing fluid indicates values for the sample channel alone. (C) Variation in the peak wavelength for bead spectra (obtained with a sheath rate of 0.174 µL/min) correlates with variation in bead position along the x-direction of the flow cell (core width).

6. Fluorescence signals for out-of-focus particles/cells

The effective Z-axis spread of particle travel in our system was set by the microfluidic T-

cell depth of 92 µm. To test the amount of fluorescence processed by the TDI SFC system when in/out of focus, we dispensed fluorescent calibration beads diluted in water onto a glass slide, sealed the slide with a coverslip, and placed the slide within the TDI SFC system and imaged through the coverslip. We focused on a single bead via the focus system's CMOS camera (**Figure 4.1A**). We displaced the bead out of the focal plane as measured by a micrometer X-Y-Z sample stage and recorded images of the out-of-focus fluorescence (**SI Figure 4.5**). Changes in the bead's image were apparent after a 30-µm displacement from the focal plane. After a 50-µm displacement, which is beyond the maximum deviation allowed by the 92 µm deep microfluidic focusing channel, the image was further distorted with signal being dispersed onto a larger area of the CCD. These data illustrate that while 1D focusing does not constrain particle position in the Z-direction, particles may still be detected by the TDI SFC system, albeit with higher variation in signal intensity. In addition, variation in the flow velocity along the z-axis (see **SI Figure 4.4A**) will also cause some particles/cells to not be synchronized to the CCD clocking rate.



SI Figure 4.5 (A) Images from the CMOS focus camera of a single fluorescent bead translated out of the TDI-SFC system's focal plane using a micrometer positioning stage. (B) Line fluorescence profiles across the bead showing that the bead when shifted out of the system's focal plane generates out-of-focus fluorescence that may still be detected by the TDI-SFC system.

4.6 References

(1) Roederer, M. *Cytometry A* **2008**, *73*, 384-385.

(2) Surre, J.; Saint-Ruf, C.; Collin, V.; Orenga, S.; Ramjeet, M.; Matic, I. *Sci Rep* **2018**, *8*, 12088.

(3) Futamura, K.; Sekino, M.; Hata, A.; Ikebuchi, R.; Nakanishi, Y.; Egawa, G.; Kabashima, K.; Watanabe, T.; Furuki, M.; Tomura, M. *Cytometry A* **2015**, *87*, 830-842.

(4) Basiji, D.; Ortyn, W. E.; George, T. C.; Brawley, J.; Hall, B. E.; Morrissey, P. *Cytom Part A* **2004**, *59a*, 87-87.

(5) Basiji, D. A.; Ortyn, W. E.; Liang, L.; Venkatachalam, V.; Morrissey, P. *Clin Lab Med* **2007**, *27*, 653-670, viii.

(6) Brawley, J.; George, T. C.; Hall, B. E.; Frost, K.; Zimmerman, C. A.; Seo, M.; Basiji, D.; Ortyn, W. E.; Morrissey, P. *Cytom Part A* **2004**, *59a*, 130-130.

(7) George, T. C.; Hall, B. E.; Zimmerman, C. A.; Frost, K.; Seo, M.; Ortyn, W. E.; Basiji, D.; Morrissey, P. *Cytom Part A* **2004**, *59a*, 118-118.

(8) United States, Patent 6249341, 2001

(9) Nolan, J. P.; Condello, D. *Curr Protoc Cytom* **2013**, *Chapter 1*, Unit1 27.

(10) Watson, D. A.; Gaskill, D. F.; Brown, L. O.; Doorn, S. K.; Nolan, J. P. *Cytometry A* **2009**, *75*, 460-464.

(11) Telford, W. G.; Shcherbakova, D. M.; Buschke, D.; Hawley, T. S.; Verkhusha, V. V. *Plos One* **2015**, *10*.

(12) Hauer, P.; Le Ru, E. C.; Willmott, G. R. *Biomicrofluidics* **2015**, *9*.

(13) Goddard, G.; Martin, J. C.; Naivar, M.; Goodwin, P. M.; Graves, S. W.; Habbersett, R.; Nolan, J. P.; Jett, J. H. *Cytometry A* **2006**, *69*, 842-851.

(14) Jackson, J. M.; Witek, M. A.; Kamande, J. W.; Soper, S. A. *Chem Soc Rev* **2017**, *46*, 4245-4280.

(15) Emory, J. M.; Soper, S. A. *Analytical Chemistry* **2008**, *80*, 3897-3903.

CHAPTER V

Application of Spectral Flow Cytometry with Time-Delayed Integration (TDI SFC)

5.1 Introduction

As a demonstration of the TDI SFC system, we undertook the immunophenotyping of CLCs from patients with B-cell acute lymphoblastic leukemia (ALL). In acute leukemias, patients in remission may harbor drug resistance, and residual disease that can progress from low measurable residual disease (MRD) to lethal levels in weeks to months.[1,2] Due to this rapid progression, frequent MRD surveillance is paramount to tailor patient-specific treatment strategies.[2,3] MFC analysis is commonly implemented to monitor B-ALL MRD using bone marrow aspirates[4], where leukemic cells are differentiated by abnormal expression of hematopoietic proteins.[5] MFC can detect marrow-derived leukemic cells at a frequency of 10^{-3} – 10^{-4}, but this sensitivity does not extend to peripheral blood due to hemodilution of CLCs.[6] Thus, standard-of-care MRD monitoring is limited by invasive bone marrow biopsies due to the high occurrence of leukemia cells in the bone marrow, and as such, monitoring for MRD can occur every few months. However, rapid relapse events associated with acute leukemias require weekly samplings to begin therapeutic intervention early to potentially provide better patient outcome.[2]

We recently demonstrated a microfluidic assay[3] to affinity-enrich CLCs from the peripheral blood of patients with acute myeloid leukemia (AML). The surfaces of microfluidic devices were coated with antibodies that selected myeloid cells from blood, and CLCs were identified by immunostaining against patient-specific leukemic antigens.[3] Depending on the

patient's disease state, total cell counts (10-100,000 cells/sample) and CLC frequencies (0-98%) were detected.[3] We have observed similar trends in an ongoing project for B-ALL patients in which CLC enrichment from blood was enabled by using microfluidic devices coated with anti-CD19 antibodies; CLCs were differentiated from normal CD19(+) cells by immunostaining against the nuclear protein TdT (terminal deoxynucleotidyl transferase).[7] Given the sample-limited cell counts, we released cells from the microfluidic device[8] for immunophenotyping by semi-automated fluorescence microscopy. While reliable, this process required lengthy manual cell counting.[3] Unfortunately, in many cases the CLC levels were below what could be accurately processed using MFC.

Therefore, we developed a spectral flow cytometer with time delayed integration mode (TDI SFC). The TDI SFC system was characterized using cell models, fluorescent calibration beads, in the previous chapter. In this chapter, we applied the system to the proof-of-concept immunophenotyping of a B-ALL cell line labeled against several markers and positively identified these cells. These data support the great potential of the TDI SFC for automated immunophenotyping of rare cells in diverse applications, including the longitudinal surveillance of residual disease in leukemias[3] and epithelial cancers.[9]

We also developed and optimized unmixing algorithms for spectral deconvolution for our TDI SFC. Spectral deconvolution ablates the need for extra samples and experiments for spectra compensation, giving rise to our TDI SFC's great potential for sample-limited CLC immunophenotyping and disease monitoring.

5.2 Experimental Methods

5.2.1 Reagents and Materials

The materials for microfluidic chip fabrications were the same as the previous chapter.

Reagents included: methanol-stabilized formaldehyde (37%), methanol-free formaldehyde (16%), phosphate buffered saline (1X PBS, pH=7.4) LIVE/DEAD™ Cell Imaging Kit (488/570), 7-aminoactinomycin D (7-AAD), and InSpeck™ Green fluorescent calibration kit from Thermo Fisher Scientific; FITC-labeled, mouse anti-human terminal deoxynucleotidyl transferase (anti-TdT) antibody from BioRad (cocktail of clones HT-1, HT-4, HT-8, HT-9); sucrose from Fluka; Histopaque®-1119, Triton™ X-100, Micro-90®, reagent-grade isopropanol (IPA) from Sigma-Aldrich; and anti-mouse Fc-specific Antibody Binding Capacity (ABC) beads from Bangs Lab. The cell filters for removing cell clumps and hemocytometer for cell counting were purchased from Thermo Fisher Scientific.

5.2.2 TDI SFC Data Processing

Acquired spectral data was collected and processed in MATLAB as described in previous chapter. Frames without any event present were eliminated based on a signal-to-noise threshold of 20 (fluorescent beads) or 10 (cell samples). The positive events were analyzed to extract event properties regarding time duration and fluorescence intensity. The frames in each positive event were integrated to obtain the total emission spectrum of. For multiplexed staining, total emission spectra were deconvoluted using classic least squares unmixing[10] from spectra acquired from singly-stained cells and autofluorescence in unstained cells.

5.2.3 Microfluidic Flow Cell Fabrication and Fluidic Connections

The 1D focusing microfluidic devices were fabricated using the same processes as described in the previous chapter. The fabrication, thermal fusion bonding and channel dimension characterization were the same as the previous chapter and were detailed in the SI of the previous chapter.

5.2.4 Cell Culturing, Immunostaining and Analysis

SUP-B15 ALL cells (ATCC CRL-1929) were cultured at 37 °C, 5% CO_2 in Iscove's Modified Dulbecco's Medium with 20% fetal bovine serum (Gibco Laboratories) and 0.02 μL 2-mercaptoethanol (MP biomedicals) per mL media. Cells were washed with PBS (300×g, 7 min) twice before and after staining processes, followed by resuspension in Histopaque®-1199. For nuclear staining, cells were fixed with 2% formaldehyde (15 min), permeabilized with 0.1% Triton™ X-100 (10 min), and stained with anti-TdT-FITC (10 μL/10^6 cells) and/or 7-AAD (4 μM) for 40 min. Unstained cells were fixed with 2% formaldehyde. For calcein staining (0.5 μM, 15 min), cells were fixed with 2% methanol-free formaldehyde to maintain membrane integrity. Successful staining was confirmed by plating cells on a glass slide and imaging with a Zeiss Axiovert 200M fluorescence microscope, XBO 75 Xe arc lamp, single band FITC (Ex: 470/20, Em: 535/40) or Cy3 (Ex: 540/25, Em: 605/55) filter set (Omega Optical), Cascade:1K EM-CCD camera (Photometrics), and MAC 5000 stage (Ludl Electronics Products). The cells were filtered with a 30 μm filter and the cell concentration was determined using a hemocytometer before introducing them to the TDI SFC with a 20x objective.

5.3 Results and Discussion

5.3.1 Detection of Leukemic Cells with Calcein Stain

The first cell sample tested by the TDI SFC were SUP-B15 ALL cells stained with calcein. Calcein acetoxymethyl (calcein AM) is a frequently used cell viability reagent that is nonfluorescent. However, calcein AM can be transported through live cell surface membrane to the intracellular matrix and generate fluorescent calcein molecules with the presence of esterases. The detachment of acetoxymethyl (AM) groups under the catalysis of esterase makes the calcein molecules fluorescent, which shows the highest emission at 516 nm with optimal excitation

efficiency at 495 nm, allowing for the cell viability information. Calcein stain in live cells is known for its bright green fluorescence signal. Since our TDI SFC adopts a 488 nm laser light source which can excite calcein molecules with high efficiency, calcein-stained leukemic cells were chosen to be tested by the system as an initial validation of application. A critical fact about calcein stain is that the cell membrane must be intact to keep the fluorescent calcein molecules inside. If the cell membrane is compromised, the calcein molecules will not be confined in the cytoplasms and will lead to failed staining process. **Figure 5.1** shows the microscopy data and TDI SFC data from this cell population.

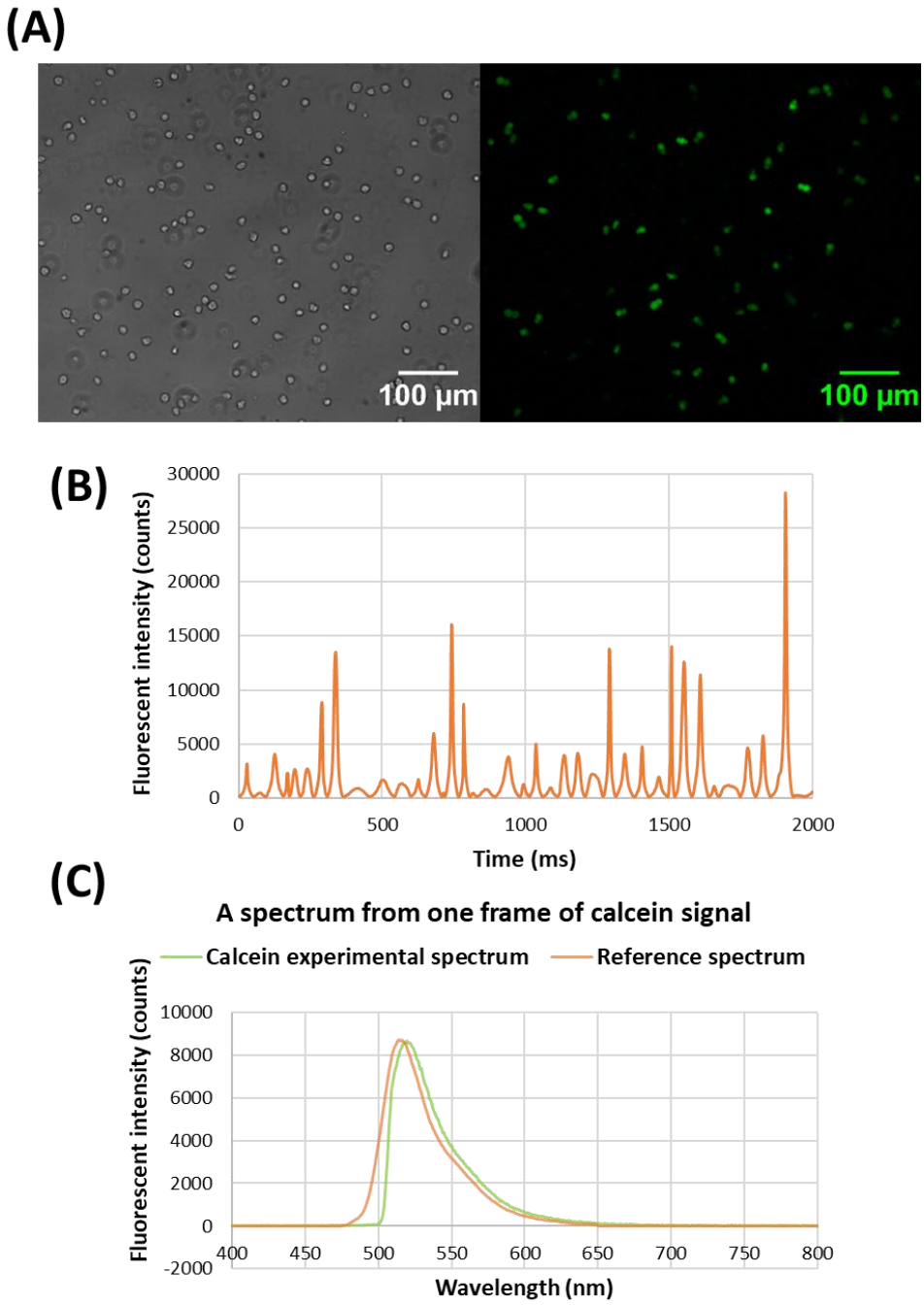


Figure 5.1 Detection of the calcein stained leukemic cells. (A) Left: a bright field image of the cells. Right: the same cells imaged in FITC channel to show the calcein signal. (B) Plot at the maximum emission (516 nm) cross the calcein-stained cell events detected by our TDI SFC. (C) A spectrum recorded by one frame from a cell event (*green*) showed agreement with the reference emission spectrum of calcein (*orange*).

The bright field image of some calcein AM treated cells and the same cell image in fluorescent channel are shown in **Figure 5.1A**. The bright field image displays all of the cells in

the field of view while the fluorescent image only shows the cells with successful staining. From the images it can be concluded that not all of the cells were successfully stained with calcein. This is mainly because of two facts. First, some of those cells were dead cells and had compromised cell surface by the time of staining. One reason attributing to this is apoptosis. Apoptosis is a genetically controlled process that results in programmed cell death to remove cells, mainly aged or unhealthy cells, to maintain appropriate cell numbers. [11] When a cell is apoptotic, although the cell can be clearly observed in the bright field image, the surface membrane integrity cannot be guaranteed, resulting in failed calcein staining. Second, although the healthy live cells were stained well when being treated with calcein AM, in the later sample preparation processes of spinning down, vortexing and pipetting to remove excess calcein AM molecules, the cell surface of some cells were ruptured due to mechanical agitation, causing release of celcein molecules from the cytoplasms.

Despite the fact that not all of the cells in the bright field image were stained with calcein, the successfully stained cells were able to be detected by the TDI SFC. The plot at the maximum emission (around 516 nm) across some representative positive events after background subtraction is shown in **Figure 5.2B**. The peaks in this figure indicates the positive calcein cell events. To validate the reliability of the data obtained from the TDI SFC, we also compared the experimental spectrum with the reference spectrum of calcein. As revealed by **Figure 5.2C**, both their maximum emission and peak shape showed good agreement with each other. The detailed TDI SFC data showed a maximum emission of 519.6 nm which was very close to the expected maximum emission of calcein, 516 nm. These preliminary data demonstrated the reliability of our TDI SFC system.

After successfully detecting the bright calcein-stained cells, to further validate the application of the TDI SFC, the same cell lines with weaker stains were tested. SUP-B15 B-ALL cells were stained with 7-AAD and anti-TdT FITC and tested by the TDI SFC using the same instrument parameters as calcein population. 7-AAD as a nuclear stain was used to make sure that the event detected was truly from a cell but not from any interferent. TdT is a nucleus enzyme that is positively expressed most ALL patients. [12] Thus, it was used as another detection marker.

5.3.2 Multiplexed detection of B-ALL CLCs

We are currently participating in a pilot clinical study to detect relapse from MRD in B-ALL patients by monitoring the number of CLCs in peripheral blood using microfluidic enrichment of CD19(+) B-cells. The enriched B-cells are immunostained to specifically identify CLCs. For this clinical application, we must differentiate CD19(+) normal cells from CD19(+) CLCs. To accomplish this, immunophenotyping is employed and those that stain positive for TdT are classified as CLCs. We chose 7-AAD as the nuclear stain for the TDI SFC system because of its large Stokes shift (λ_{max} = 647 nm), which enables multiplexing with dyes such as FITC, PE, or tandem dyes. We evaluated the multiplexing capabilities of the TDI SFC by analyzing mixtures of FITC and streptavidin conjugated with PE, PE-Cy5, PE-Cy7, or PerCP-Cy5.5 using the TDI SFC. Least squares unmixing[10] was used to deconvolute the resultant spectra into the individual components and we found that we could correctly identify 11 different dye mixtures (**SI Figure 5.1**).

As surrogates for CLCs, we stained a model B-ALL cell line (SUP-B15) that expressed TdT. For the cell enumeration we used 7-AAD as the nuclear stain, and/ FITC-conjugated anti-TdT Abs to check for TdT expression. In this cell line, we have consistently observed strong

autofluorescence in ~6% of the SUP-B15 cells throughout normal culture conditions (**Figure 5.3)**, which were likely a subpopulation of apoptotic cells.[13] We differentiated these cells from positive FITC-immunostaining of a membrane protein by morphological examination, where immunostaining was exclusive to the membrane whereas autofluorescence was observed throughout the cell.[3] Rather than morphological examination by microscopy or imaging flow cytometry,[14] TDI SFC uses high spectral resolution to differentiate FITC's or 7-AAD's emission from autofluorescence.

To evaluate the spectral properties of the fluorescent components, we processed SUP-B15 cells that were fixed and permeabilized and either unstained (autofluorescence spectrum) or stained with 7-AAD and/or anti-TdT-FITC Abs. The experimental TDI SFC emission spectra matched well with reference spectra for the singly-stained controls (**Figure 5.2**) and served as a training set for multiplexed spectral deconvolution.

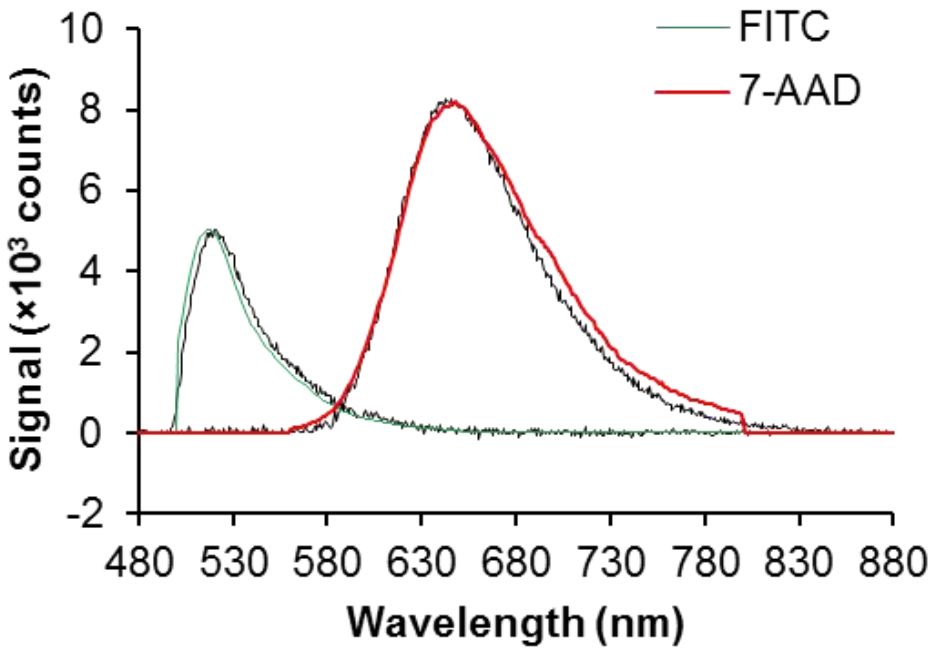


Figure 5.2 Sample TDI SFC spectra (black lines) for FITC and 7-AAD compared to reference spectra (colored lines)

5.3.3 Autofluorescence in SUP-B15 B-ALL cell line

We observed autofluorescence in a subpopulation of SUP-B15 B-ALL cells that persisted throughout normal culture conditions. In **Figure 5.3**, we show signal in the FITC channel by fluorescence microscopy for a fraction of the SUP-B15 cells that we attributed to autofluorescence with similar autofluorescence appearing in the Cy3 and Cy5 channels (data not shown). To accurately quantitate how many SUP-B15 cells were autofluorescent, we interrogated unstained cells by traditional flow cytometry (**Figure 5.3B**). We observed two subpopulations of cells based on the forward scattering (FSC) signal. Approximately half of the cells generated twice as much FSC signal as the other half, concordant with microscopic examinations (**Figure 5.3A**). We suspect these cells were actively mitotic and therefore larger.

By examining signal in the FITC channel (**Figure 5.3C**), we observed that only a subpopulation of the smaller cells (lower FSC) exhibited increased autofluorescence (~6.4% of all cells in the FSC/SSC gate in **Figure 5.3B**). Given the lack of increased autofluorescence in larger cells suspected to be mitotic, we hypothesize that roughly 6% of SUP-B15 cells are apoptotic. We did not test this hypothesis due to interference of autofluorescence with standard annexin V fluorescence staining. However, there is substantial literature precedence for increased autofluorescence at 488 nm during apoptosis, particularly due to oxidation of mitochondrial flavins, such as flavin adenine dinucleotide (FAD).[13]

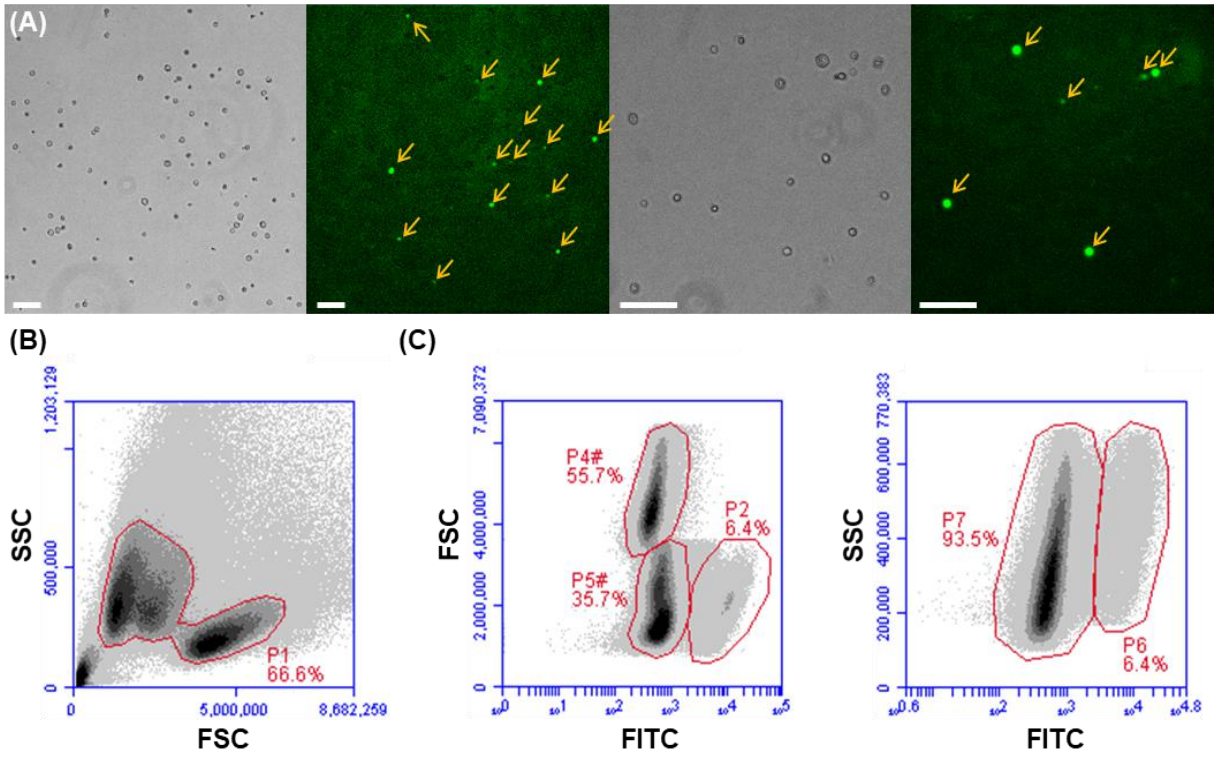


Figure 5.3 (A) Brightfield (grey) and fluorescence (FITC channel, green) microscopy images of unstained SUP-B15 cells. Yellow arrows mark the presence of cells with increased autofluorescence. All scale bars equal 100 µm. (B) Flow cytometry of unstained cells showed two subpopulations (outlined by red) of cells with the same side scattering (SSC) signal but differing forward scattering (FSC) signal. (C) Only the smaller cells (~6.4% of total cells) with lower FSC signal exhibited increased autofluorescence on the FITC channel, whereas we found no correlation between increased FITC signal and SSC. We also note that we observed similar trends on all fluorescence channels (PE, PerCP, APC; data not shown)

5.3.4 Deconvolution of Multiplexed Fluorescence Spectra from B-ALL Cells

Classic least squares unmixing[10] was unable to compensate for slight shifting of the emission spectra with variation in the cells' x-axis lateral position (Chapter 4 **SI Figure 4.4C**). Thus, we fit each emission spectrum with multiple Gaussian functions and used a nonlinear least squares solver with strict constraints (**Figure 5.4A, Table 5.1**) of the Gaussians' relative wavelength positions, standard deviations, and intensities. We note that singly-stained controls need not be repeated as the fluorescent dye's spectral properties remain constant.

After processing the cells with the TDI SFC system, we background subtracted each event and summed signals in all frames throughout a single event duration. These integrated spectra were deconvoluted by curve fitting to determine the contribution of each fluorophore component. Each curve, F(λ), was represented via sums of Gaussian functions:

$$F(\lambda) = \sum_{f=1}^{3} I^f \sum_{i=1}^{N^f} A_i^f e^{-\left(\frac{x-\left(\lambda_1^f+\Delta\lambda_i^f\right)}{\sigma_i^f}\right)^2} \quad \text{(Eq. 5.1)}$$

In Eq. (5.1), I^f is the intensity of fluorophore f; N^f is the number of Gaussian functions for fluorophore f; and A_i^f is the relative intensity of fluorophore f's i^{th} Gaussian function, which is defined by the standard deviation in wavelength, σ_i^f, and the Gaussian's peak position, $\Delta\lambda_i^f$, that is set relative to the first Gaussian function, λ_1^f.

We used the unstained or singly-stained SUP-B15 cells to determine the curve fitting parameters for autofluorescence and the 7-AAD and FITC fluorophores. We hand selected the 7-AAD and FITC training sets to be reasonably devoid of autofluorescence components. These spectra agreed well with reference spectra (Thermo Fisher Fluorescence SpectraViewer, **Figure 5.2**).

We note that slight variation in the cells' lateral positions translated to shifting of the fluorescence spectra along the CCD's wavelength axis, generally by 1-5 nm. Thus, each fluorophore's components were not hard set but allowed to slightly shift in position and magnitude according to constraints within a nonlinear, least squares solver. We outline the curve fitting constraints as well as the median values observed in the training sets in **Table 5.1** and show the residuals, which were normalized to the maximum signal and averaged over every event in the training set (**Figure 5.4B**). Lastly, we note that nonlinear least squares solvers are highly susceptible to poor initial conditions for sums of exponentials. For multiplexed

deconvolution, we first used the median values of fitting constraints in **Table 5.1** to generate an initial condition for a solution using variable fitting constraints.

Table 5.1 Curve fitting constraints for FITC, 7-AAD, and autofluorescence spectra that were optimized via training sets for a nonlinear, least squares solver.

Component (events)	Variable	Fitting constraints		Median values	
FITC (N=33)	A_1	1.00-1.00		1.00	
	λ_1	522-526	nm	523.7	nm
	σ_1	8-11	nm	9.6	nm
	A_2	1.00-1.00		1.00	
	$\Delta\lambda_2$	11-14	nm	12.2	nm
	σ_2	11-17	nm	14	nm
	A_3	0.57-0.64		0.62	
	$\Delta\lambda_3$	28-35	nm	31.4	nm
	σ_3	21-27	nm	22.9	nm
	A_4	0.18-0.23		0.21	
	$\Delta\lambda_4$	60-66	nm	64	nm
	σ_4	32-38	nm	38	nm
	A_5	0.40-0.48		0.44	
	$\Delta\lambda_5$	-10,-8	nm	-8.9	nm
	σ_5	4-7	nm	5.3	nm
7-AAD (N=165)	A_1	1.00-1.00		1.00	
	λ_1	637-650	nm	638.8	nm
	σ_1	24-34	nm	28.5	nm
	A_2	1.00-1.00		1.00	
	λ_2	25-35	nm	28.6	nm
	σ_2	33-43	nm	32.4	nm
	A_3	0.34-0.40		0.18	
	λ_3	65-71	nm	94	nm
	σ_3	30-36	nm	63	nm
	A_4	0.15-0.21		0.37	
	λ_4	88-94	nm	66.9	nm
	σ_4	57-63	nm	33.2	nm
Autofluorescence (N=22)	A_1	0.30-0.34		0.33	
	λ_1	518-521	nm	518.9	nm
	σ_1	6-8	nm	6.2	nm
	A_2	0.54-0.58		0.57	
	$\Delta\lambda_2$	12-15	nm	13.4	nm
	σ_2	12-15	nm	13.5	nm
	A_3	0.78-0.82		0.64	
	$\Delta\lambda_3$	31-34	nm	97.9	nm
	σ_3	19-23	nm	50.2	nm
	A_4	0.85-0.90		0.27	
	$\Delta\lambda_4$	61-63	nm	146	nm
	σ_4	33-36	nm	77.3	nm

A_5	0.62-0.66		0.81	
$\Delta\lambda_5$	96-99	nm	33.7	nm
σ_5	48-51	nm	22.4	nm
A_6	0.27-0.31		0.90	
$\Delta\lambda_6$	142-146	nm	62.4	nm
σ_6	74-79	nm	34.6	nm

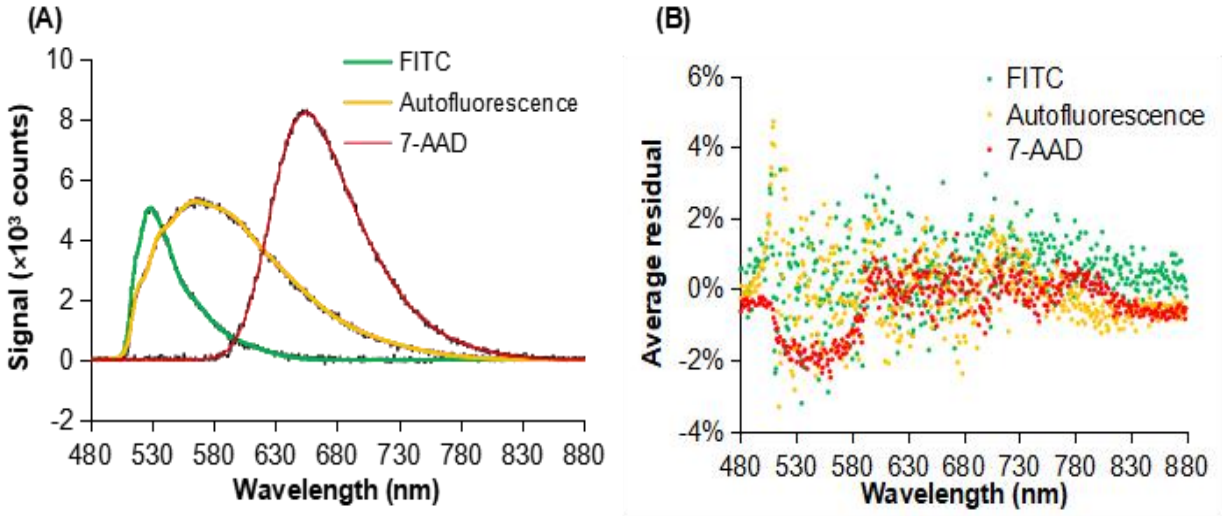


Figure 5.4 (A) Sample spectra and fitting for FITC, 7-AAD, and autofluorescence, where the raw data is shown in black beneath the fit curve. (B) After fitting all curves in the training set using the constraints in **Table 5.1**, the average residual for the curve fitting was computed and normalized to the maximum signal in the raw data.

We excluded highly autofluorescent SUB-15 cells (**SI Figure 4.2**) and differentiated normal B-cells (cells stained positive for 7-AAD only) from CLCs (cells stained positive for both 7-AAD and anti-TdT-FITC Abs). By multiplexed deconvolution, TDI SFC identified 100% of cells as singly-stained and correctly identified 99.1% of the CLCs as positive for both markers using the TDI SFC system (**Figure 5.5 A**). We observed slightly higher median S_{int} signals for 7-AAD (1.7×10^5 counts) compared to anti-TdT-FITC Abs (1.1×10^5 counts), but equivalent SNR (175) for the stains (**Figure 5.5 B**).

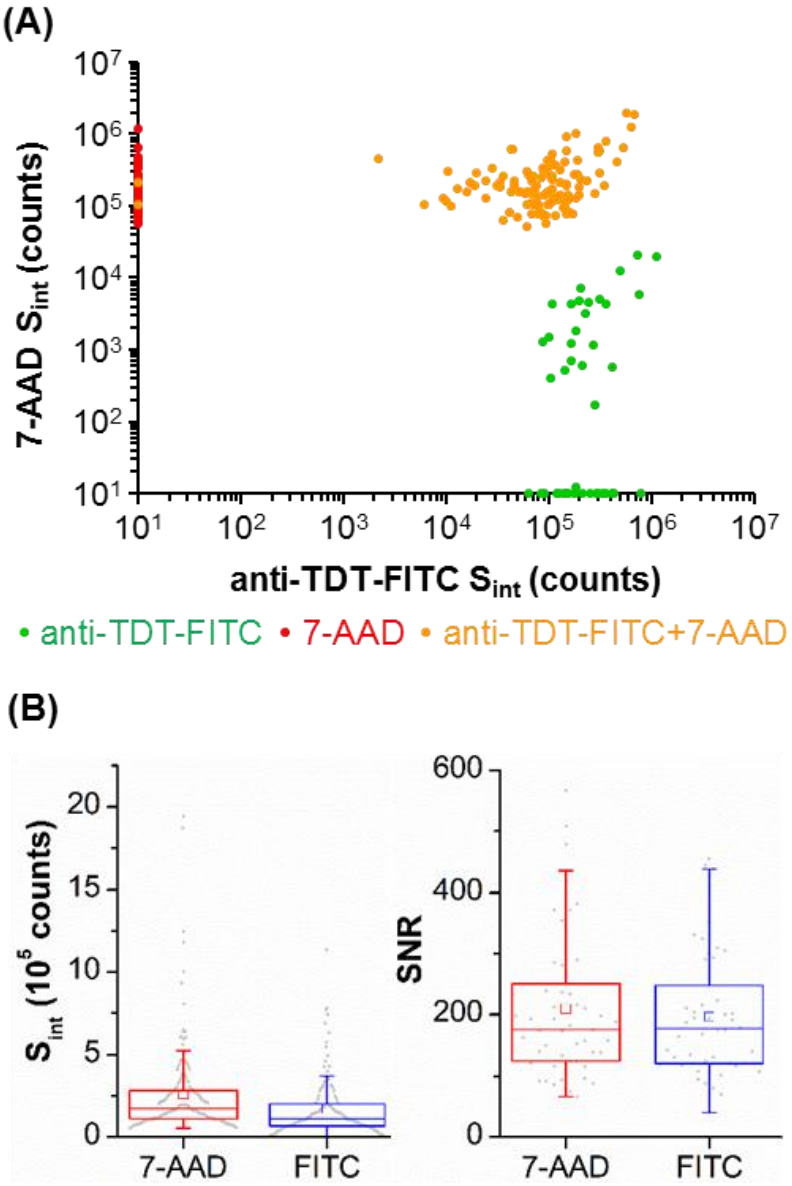


Figure 5.5 (A) Multiplexed deconvolution correctly classified 100% of singly-stained cells and 99.1% of dual-stained model CLCs. (B) Box plots of S_{int} and SNR for 7-AAD or anti-TdT-FITC staining.

TDI SFC classification was based on each cell's unique fluorescence emission spectrum. This is in contrast to conventional flow cytometry, where emission is spectrally binned and cells with similar emission intensities are clustered for identification. This unique aspect of TDI SFC spectral deconvolution enabled us to achieve high accuracy in classifying cells even at low cell numbers (46-111 cells per group) that would normally be difficult for traditional flow cytometry.

5.4 Conclusions

As a demonstration of the utility of the TDI SFC, we processed B-ALL cells stained with a nuclear dye (7-AAD) and a leukemic marker (anti-TdT-FITC). A ~100% classification accuracy was secured for the CLCs by deconvoluting TDI SFC spectra for each cell, which were readout with high signal (10^6 integrated counts) and SNR (175), at cell numbers (46-111 cells) that would be prohibitive for traditional MFC. We provided further evidence for the spectral deconvolution of 5 multiplexed dyes and compared TDI SFC performance to a commercial MFC instrument (97% concordance for cell detection efficiency).

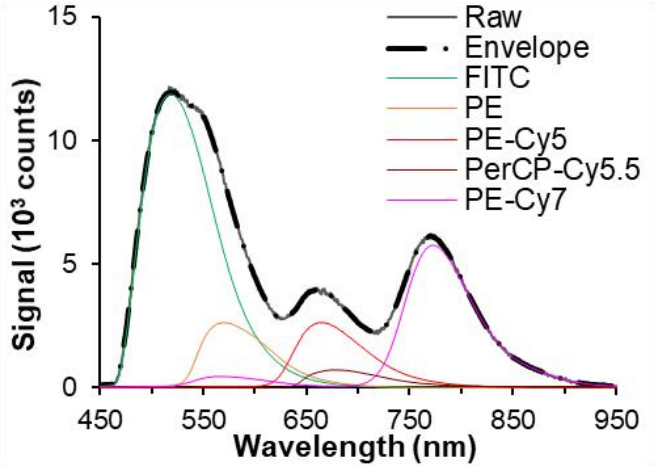
Limitations of the current TDI SFC instrument was the relatively modest throughput; 90 events/min. This was a result of the limited readout rate of the CCD used herein; lower serial register readout times can improve this. Also, in the current format of the TDI SFC, the CV was 27.8% compared to 5% for the commercial MFC. This was due to the 1D microfluidic focusing employed herein. Future renditions of this TDI SFC instrument will seek to use 2-D microfluidic focusing to ensure that all cells are focused to the focal plane of the epi-illumination objective and proceed through the field-of-view with better velocity uniformity and fluorescence collection efficiency.

5.5 Appendix-Supplemental Information

Table of Contents

1. Classic least squares unmixing of dye mixtures

We infused mixtures of FITC (1 nM), streptavidin-PE (5 nM), streptavidin-PE-Cy5 (5 nM), and/or streptavidin-PE-Cy7 (50 nM) through a glass capillary. For every possible combination of dyes, results from classic least squares unmixing[10] correctly identified each mixture's composition (**SI Figure 5.1**) without false positives or negatives.

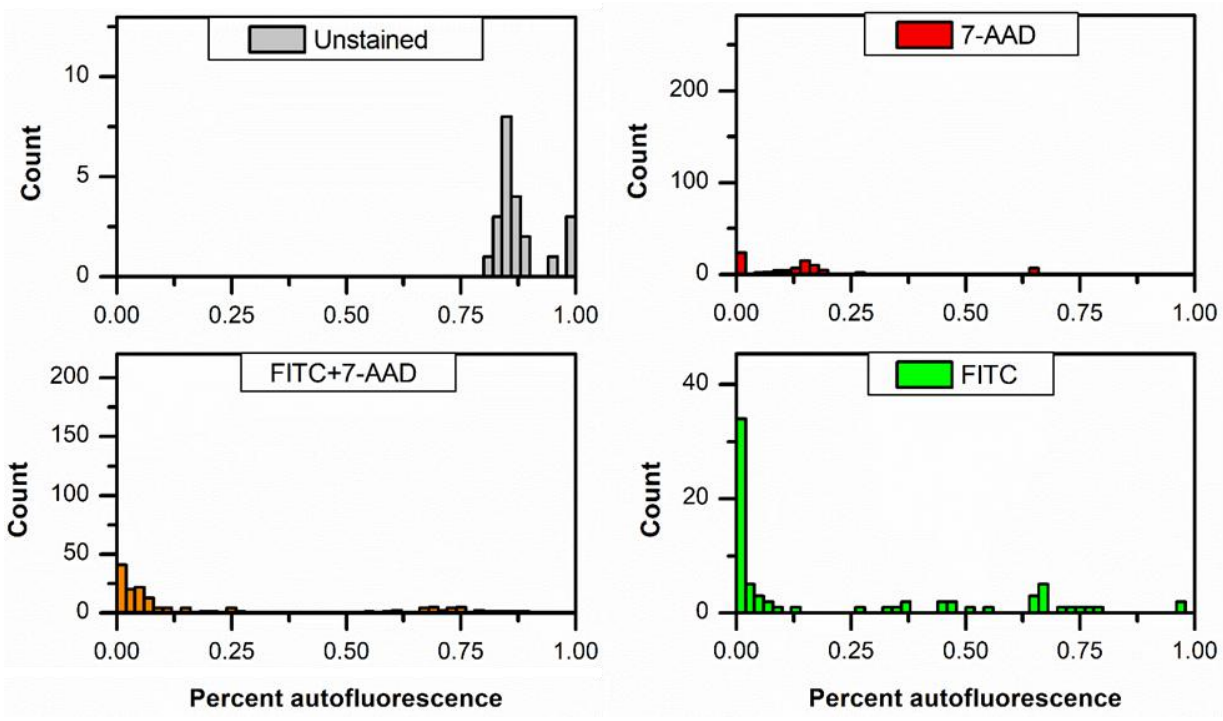


Dyes detected λ_{em}	Composition of mixture tested: FP	F5	F7	FP5	FP7	F57	FP57	P5	P7	P57	57
FITC (F) 518 nm	F	F	F	F	F	F	F	-	-	-	-
PE (P) 575 nm	P	-	-	P	P	-	P	P	P	P	-
PE-Cy5 (5) 670 nm	-	5	-	5	-	5	5	5	-	5	5
PE-Cy7 (7) 767 nm	-	-	7	-	7	7	7	-	7	7	7

SI Figure 5.1 Mixtures of FITC (F), PE (P), PE-Cy5 (5), and PE-Cy7 (7) were imaged using the TDI-SFC system. Along the columns, mixtures are designated by their components, where FP includes FITC and PE but not PE-Cy5 or PE-Cy7 and FP57 includes all components. Resultant spectra were deconvoluted by classic least squares unmixing, and the dyes detected for each mixture are shown along the rows. If a dye was detected in the mixture, the cell is colored and labeled, whereas if the dye was not detected by deconvolution, the cell is grey and marked with a dash. For every mixture tested, the system correctly identified the dye components without any false positives.

2. Autofluorescence in SUP-B15 B-ALL cell line

SUP-B15 cells were autofluorescent. Autofluorescence spectra overlapped significantly with both FITC and 7-AAD spectra (**Figure 5.4A**), so we sought to eliminate interfering autofluorescence by deconvolution. We examined the percentage of autofluorescence signal in each event (**SI Figure 5.2**). All unstained cells had little signal attributed to either FITC or 7-AAD by deconvolution, whereas most stained cells had <25% autofluorescence. Thus, in subsequent analysis we excluded cells with ≥25% autofluorescence signal.



SI Figure 5.2 Percentage of signal originating from autofluorescence by multiplexed spectral deconvolution. In all subsequent analyses, cells with ≥25% autofluorescence were excluded.

5.6 References

(1) Ryan, D. H.; Dongen, J. J. M. v. **1988**.

(2) Schuurhuis, G. J.; Ossenkoppele, G. *Expert review of hematology* **2010**, *3*, 1-5.

(3) Jackson, J. M.; Taylor, J. B.; Witek, M. A.; Hunsucker, S. A.; Waugh, J. P.; Fedoriw, Y.; Shea, T. C.; Soper, S. A.; Armistead, P. M. *Analyst* **2016**, *141*, 640-651.

(4) Wood, B. L. *Cytom Part B-Clin Cy* **2016**, *90*, 47-53.

(5) Wood, B. L. *Journal of Hematopathology* **2015**, *8*, 191-199.

(6) Paietta, E. *Hematology / the Education Program of the American Society of Hematology. American Society of Hematology. Education Program* **2012**, *2012*, 35-42.

(7) Seegmiller, A. C.; Kroft, S. H.; Karandikar, N. J.; McKenna, R. W. *Am J Clin Pathol* **2009**, *132*, 940-949.

(8) Nair, S. V.; Witek, M. A.; Jackson, J. M.; Lindell, M. A. M.; Hunsucker, S. A.; Sapp, T.; Perry, E.; Hupert, M. L.; Bae-Jump, V.; Gehrig, P. A.; Wysham, W. Z.; Armistead, P. M.; Voorhees, P.; Soper, S. A. *Chem Commun* **2015**, *51*, 3266-3269.

(9) Jackson, J. M.; Witek, M. A.; Kamande, J. W.; Soper, S. A. *Chem Soc Rev* **2017**, *46*, 4245-4280.

(10) Nolan, J. P.; Condello, D. *Curr Protoc Cytom* **2013**, *Chapter 1*, Unit1 27.

(11) Elmore, S. *Toxicologic pathology* **2007**, *35*, 495-516.

(12) Bassan, R.; Gatta, G.; Tondini, C.; Willemze, R. *Critical Reviews in Oncology Hematology* **2004**, *50*, 223-261.

(13) Surre, J.; Saint-Ruf, C.; Collin, V.; Orenga, S.; Ramjeet, M.; Matic, I. *Sci Rep* **2018**, *8*, 12088.

(14) Basiji, D. A.; Ortyn, W. E.; Liang, L.; Venkatachalam, V.; Morrissey, P. *Clin Lab Med* **2007**, *27*, 653-670, viii.

Chapter VI

Conclusions and Future Directions

6.1 Summary and Conclusion

Leukemia is a serious health concern that necessitates more sensitive methods for monitoring rare MRD in PB instead of invasive BM biopsies. Our group previously developed a microfluidic assay for CLC monitoring in AML patients' PB samples with a sensitivity of 10^{-5}-10^{-6} compared with a sensitivity of 1-5%[1] with morphology and 10^{-3}-10^{-4} with traditional flow cytometry.[2] The chip-isolated cells in this microfluidic assay were enumerated and immunophenotyped by semi-automated fluorescence microscopy. But we would like to increase the automation of the immunophenotyping process. Flow cytometry is a commonly used method for cell immunophenotyping with higher throughput than fluorescence microscopy. However, because the number of cells isolated in our microfluidic assay is normally in the range of 100-10,000, these cells are not adequate for immunophenotyping in traditional flow cytometry due to the requirement of extra sample for setting up proper gattings and performing spectra compensation. Therefore, we developed a TDI SFC for chip-isolated cell enumeration and immunophenotyping with the ultimate goal of closely monitoring CLCs in acute leukemia patients, especially in B-ALL adolescent leukemia patients.

The TDI SFC has the following unique features. 1) It operates in a TDI mode. With TDI, when the cell's linear velocity is synchronized with the TDI CCD's readout rate the photoelectrical signal will be integrated in only a few pixel rows at readout, increasing SNR than the normal full frame readout.[3] 2) There is no shutter needed during TDI data acquisition because the exposure and readout occur simultaneously. This resulted in nearly 100% duty cycle, greatly reduced the possibility of missing cells.

We carefully characterized the TDI SFC with stable, uniform and bright fluorescent latex beads in several aspects. 1) We fabricated a microfluidic flow cell to hydrodynamically focus the cell sample to increase the uniformity of the cells' linear velocities. A square capillary was also used as a flow device to confirm the benefit of hydrodynamic focusing the in microfluidic flow cell. 2) We found out the optimal sample and sheath flow rates in the microfluidic flow cell to synchronize the analytes with the TDI CCD's readout. 3) Various optimizations were performed to further increase the fluorescent signal and SNR. 4) Detection efficiency and throughput were investigated to find out the upper limit of throughput of our system

After the characterizations and optimizations, we applied the TDI SFC to a proof-of-concept application to validate its potential in leukemic MRD monitoring. SUP-B15 B-ALL leukemic cells were stained with against markers and tested in the system. The TDI SFC successfully detected the cell events from each population stained with different markers. It also successfully detected some unexpected events in the unstained negative control. We suspect these events were mainly from autofluorescent cells. We also developed and optimized custom-developed spectra unmixing algorithms to perform spectra deconvolution on the TDI SFC data. By using the spectra of singly stained and unstained cell populations as training set, the algorithms successfully resolved each component spectrum in the dually stained cell population and identified 99.1% of the cells as positive for both markers. All these data demonstrated the great potential of our TDI SFC in enumerating and immunophenotyping chip-isolated CLCs from leukemia patients' peripheral blood.

6.2 Future Direction

6.2.1 2D Focusing Microfluidic chip

There are several aspects where improvement can be made to make the TDI SFC's performance better and facilitate its great potential in monitoring MRD in leukemia patients. One will be using a 2D focusing chip instead of 1D focusing chip. The current microfluidic flow cell provides 1D focusing only on the lateral axis (x axis in the schematic in **Figure 4.1**). There is no flow focusing on the z axis where the channel depth is 93 μm. As a result, the cells can be spread out at the channel depth of 93 μm, which is much more than the depth of field of the optical system.

On the object plane, depth of field Z defines the depth along the z axis where the objects of the specimen is in focus.[4] For diffraction-limited optical systems, Z can be expressed as

$$Z = n\lambda/NA^2 \quad \text{(Eq. 6.1)}$$

where n is the refractive index of the medium between the lens and the object, which is 1.0003 here; λ is the wavelength of light in air, which is 515 nm for FITC here; and NA is the numerical aperture of the objective lens. Thus, Z (10x) is 2.08 μm and Z (20x) is 0.9 μm in our TDI SFC system.

When the specimen is out of the depth of field, it is not well-focused by the optical system any more, leading to low quality image or no signal at all. Specifically in our system, if the cell is at the top or bottom of the microfluidic channel, it will be possible that only weak signal is obtained or the cell is not detected at all (**Figure 6.1A**). This is especially true for dimly expressed markers on cells. For brightly expressed markers on cells, this imperfection results in position-dependent fluorescence signal within cells that expressed exactly the same level of target markers. For two cells that express exactly the same amount of markers, the cell perfectly

within the depth of field would have more photons detected than the cell out of the depth of field, causing different signals detected by the TDI SFC system. Since this signal difference is due to the different positions of the cells but not the marker expression level difference in the cells, it should be substantially minimized. An alternative to solve this problem is to develop microfluidic flow cells capable of 2D focusing[5] (**Figure 6.1B**).

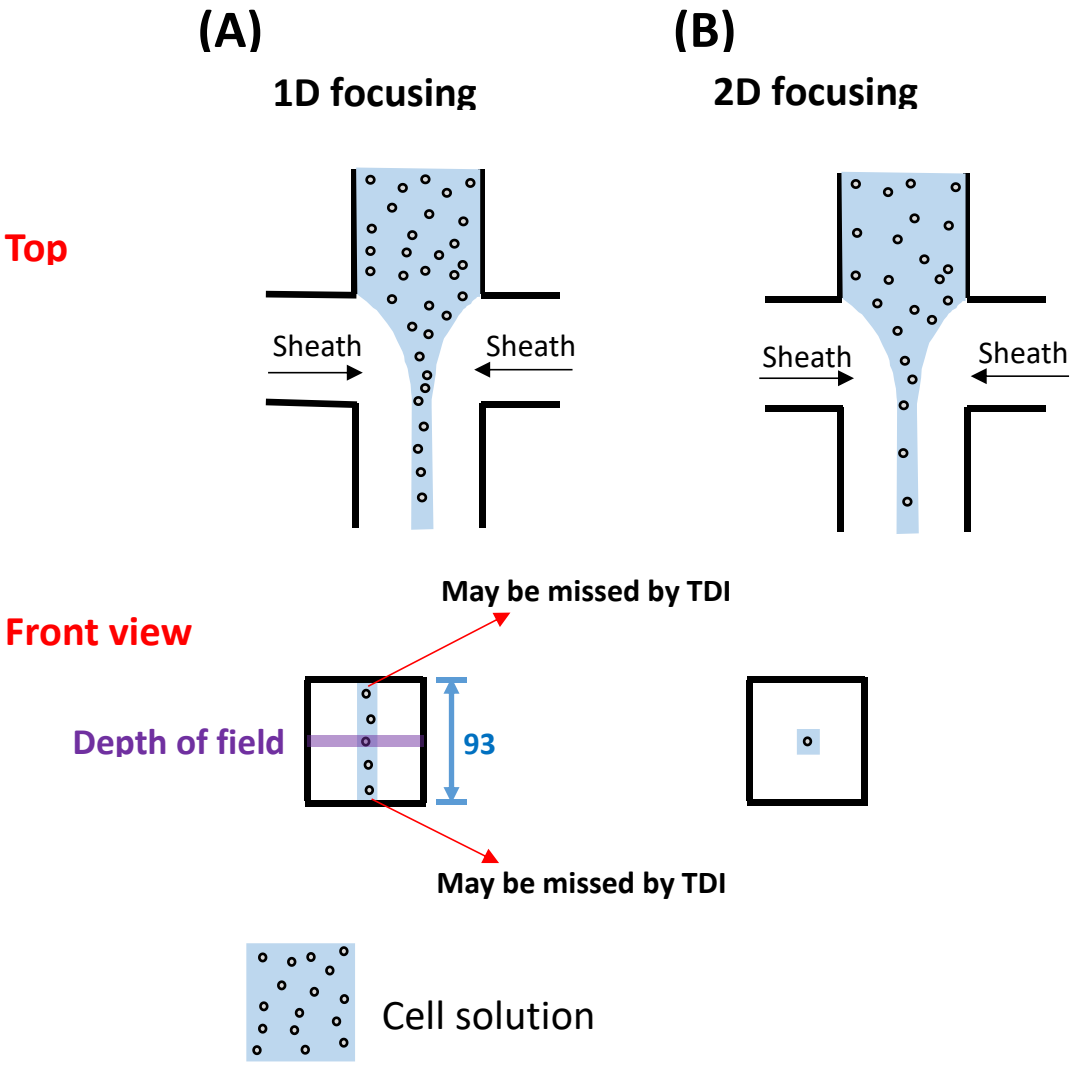


Figure 6.1 (A) 1D hydrodynamic focusing in our current microfluidic flow cell. Due to the fact that the channel depth is much more than the depth of field of the optical system, when the cells are out of the depth of field, for example, they are at the top or bottom of the microfluidic channel, it is possible that these cells are not detected by the TDI SFC. (B) An alternative to solve this problem is to develop microfluidic flow cells capable of 2D focusing.

To make our TDI SFC capable of better performance, we fabricated a mold for fabricating microfluidic flow cells that could provide 2D hydrodynamic focusing. Figure 6.2 shows the mold, a successfully bonded chip, and the schematic of 2D focusing in this chip

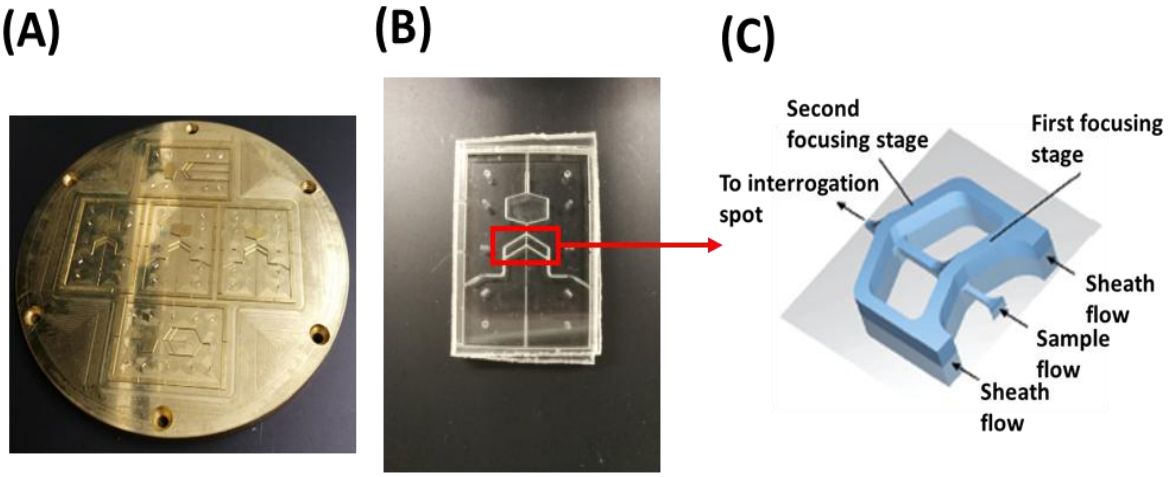


Figure 6.2 2D focusing microfluidic flow cell. (A) The brass mold for fabricating the microfluidic flow cell capable of 2D focusing. (B) A successfully bonded 2D focusing microchip. (C) The schematic of the 2D focusing in the microchip. The mold structure was transferred to both the coverslip and the substrate before chip bonding. The sheath fluid can not only focus the sample flow on the horizontal direction but also on the vertical direction, allowing for 2D focusing.

Whether this 2D focusing chip is applicable in our TDI SFC system is unknown. Tests with fluorescent calibration beads in this chip are needed to reveal whether such low flow rate necessary for the TDI CCD synchronization is applicable in this microfluidic chip. This is because when using a syringe pump to introduce sample fluid to the microfluidic chip, there is a lower limit of flow rate that can be realized due to mechanical reasons in the pump. Specifically, there is a lower limit of frequency on the pump motor to push the syringe plunger when introducing samples into the microfluidic chip. If such low flow rate necessary for current TDI synchronization is not applicable in the 2D focusing chip, one solution will be increasing the frame rate of the TDI CCD to facilitate faster flow rate of the samples.

6.2.2 Faster Frame Rate

Another modification that can improve the performance of the system is to improve the CCD camera's frame rate. If the frame rate is faster, the cell samples can move faster in the microfluidic channel. This can not only improve the throughput and make data collection more time-efficient, but also can reduce the probability of the cells' sedimentation in the fluidic system and obtain higher detection efficiency. One way to achieve this is to use binning on the pixels. For instance, if x=2, y=1 binning is used, for each pixel row there will be 670 pixels to be read out instead 1340 pixels, allowing for faster frame rate. One drawback it brings is the decreased spectral resolution. With x=2 binning, the spectral resolution is decreased from 0.7 nm/pixel to 1.4 nm/pixel. However, this decrease is not detrimental since a spectral resolution of 1.4 nm/pixel is still relatively high. We already carried out experiments on this idea. However, the readout time per pixel row decreased from about 2 ms to 1.7 ms (data not shown here). This minor decrease is not able to provide tremendous improve on frame rate. Another way to increase the frame rate will be replacing the current camera with a new one that has higher frame rate.

6.3 References

(1) Coustan-Smith, E.; Behm, F. G.; Sanchez, J.; Boyett, J. M.; Hancock, M. L.; Raimondi, S. C.; Rubnitz, J. E.; Rivera, G. K.; Sandlund, J. T.; Pui, C. H.; Campana, D. *Lancet* **1998**, *351*, 550-554.

(2) Al-Mawali, A.; Gillis, D.; Lewis, I. *American Journal of Clinical Pathology* **2009**, *131*, 16-26.

(3) Emory, J. M.; Soper, S. A. *Analytical Chemistry* **2008**, *80*, 3897-3903.

(4) Murphy, D. B. *Fundamentals of light microscopy and electronic imaging*; John Wiley & Sons, 2002.

(5) Kummrow, A.; Theisen, J.; Frankowski, M.; Tuchscheerer, A.; Yildirim, H.; Brattke, K.; Schmidt, M.; Neukammer, J. *Lab on a Chip* **2009**, *9*, 972-981.